

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
6 October 2005 (06.10.2005)

PCT

(10) International Publication Number  
**WO 2005/092918 A2**

(51) International Patent Classification<sup>7</sup>: **C07K 14/31**,  
A61K 39/085, 48/00, C12N 15/62, C07K 19/00, C12N  
5/06, 7/01

(21) International Application Number:  
PCT/GB2005/001084

(22) International Filing Date: 22 March 2005 (22.03.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0406408.5 22 March 2004 (22.03.2004) GB  
0412419.4 3 June 2004 (03.06.2004) GB

(71) Applicant (for all designated States except US): **UNI-  
VERSITY COLLEGE LONDON** [GB/GB]; Gower  
Street, London WC1E 9BT (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CHAIN, Benjamin**,  
**Michael** [GB/GB]; 46 Cleveland Street, London W1T 4JF  
(GB). **NAIR, Sean, Paddy** [GB/GB]; 256 Grays Inn Road,  
London, WC1X 8LD (GB).

(74) Agent: **WOODS, Geoffrey, Corlett**; J.A. Kemp & Co., 14  
South Square, Gray's Inn, London WC1R 5JJ (GB).

(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,  
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,  
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,  
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,  
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,  
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ,  
TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA,  
ZM, ZW.

(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,  
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,  
FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO,  
SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,  
GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: TARGETING POLYPEPTIDE

(57) Abstract: A targeting polypeptide is provided that may be used to target a chosen antigen to an antigen presenting cell. Complexes comprising such targeting polypeptide and antigen, nucleic acids and vectors encoding them, and cells comprising the nucleic acids and vectors may be used in methods of immunisation and enhance the immunogenicity of the antigen.



**WO 2005/092918 A2**

## **TARGETING POLYPEPTIDE**

### **Field of the Invention**

The present invention relates to targeting polypeptides and their use in  
5 targeting antigens to antigen presenting cells (APCs).

### **Background of the Invention**

*Staphylococcus aureus* is a major causative agent of community and hospital  
acquired infections worldwide. The organism is an important pathogen due to a  
10 combination of invasiveness, toxin production, and antibiotic resistance. *S. aureus*  
causes a wide variety of clinical syndromes, ranging from uncomplicated infections of  
the skin to life-threatening toxic shock syndrome (TSS). The bacterium causes disease  
by producing large numbers of exoproteins or virulence factors. Among the many  
known virulence factors are two families of staphylococcal pyrogenic superantigens,  
15 staphylococcal enterotoxins (SEs) and toxic shock syndrome toxin-1 (TSST-1).

Superantigens bind to major histocompatibility complex (MHC) class-II on  
antigen presenting cells (B cells, monocytes, and dendritic cells) outside the classical  
antigen-binding groove, and activate T cells by binding with the variable region of the  
 $\beta$  chain of T cell receptors ( $V\beta$ -TCR). This cross-linking triggers the non-specific  
20 activation and proliferation of T cells, induces the production of high levels of a  
variety of cytokines, and causes toxic shock syndrome characterized by fever, rash,  
hypotension and multiple organ failure. Staphylococcal enterotoxins are responsible  
for many cases of food poisoning (intoxication) associated with ingestion of toxin-  
contaminated food. To date, more than thirteen staphylococcal enterotoxins have been  
25 described.

As well as these classical superantigens, *S. aureus* also produces a family of  
proteins that have sequence homology to the superantigens, these proteins are known  
as the staphylococcal exotoxin-like proteins (SETs) and are a family of polymorphic  
paralogs. They were first identified as a genetic locus encoding at least five exotoxin-  
30 like proteins (SET1-5). More recently, data from the sequencing of the genomes of  
several different *S. aureus* strains has revealed a large number of related (36-67%) *set*  
genes clustered on a genomic island. This putative pathogenicity island, which is  
present in all strains of *S. aureus* examined to date, codes for between seven and  
fourteen *set* genes, which have varying degrees of sequence homology. In addition,

there appears to be extensive inter-strain allelic polymorphism for each of the *set* genes. The International Nomenclature Committee for Staphylococcal Superantigen Nomenclature (INCSS) has recently recommended that the SETs be renamed staphylococcal superantigen-like exoproteins (SSLs) and numbered from SSL1 to SSL14 in clockwise order from the replication origin of the chromosome based on homology to the full complement of genes found in strain MW2. This nomenclature is essentially as described by Fitzgerald *et al* (*Infect. Immun.*, 2003, 71, 2827-2838) except that the numbering of the genes is in the opposite direction. To differentiate between allelic variants the *ssl* gene is prefixed by the strain name.

The three-dimensional structure of one member of the family, SSL5 (previously SET3) has been determined. The crystal structure of this protein shows many of the characteristic structures of the superantigen superfamily, but significant differences also exist. In addition, SSLs do not show the main properties of classical superantigens, such as polyclonal T cell activation, pyrogenicity, or enhancement of endotoxin shock. The function of SSLs is therefore unknown.

### **Summary of the Invention**

The present invention is based on the finding that Staphylococcal superantigen-like exoproteins (SSLs) are able to target themselves to antigen presenting cells (APCs). This targeting to antigen presenting cells and hence the antigen presentation pathway of these cells means that a chosen antigen can also be targeted to the antigen presenting cell facilitating presentation of the antigen and hence increasing immunogenicity. The invention may also be used to target antigens to antigen presenting cells in order to induce tolerance.

Accordingly, the invention provides for the use of a complex comprising:

- (a) a targeting polypeptide comprising a staphylococcal superantigen-like protein (SSL), a fragment thereof or a variant of either, where the SSL, fragment or variant has the ability to target the complex to an antigen presenting cell; and
- (b) an antigen and/or a nucleic acid molecule encoding an antigen,

in the manufacture of a medicament for use in immunization or the induction of tolerance.

The invention also provides a complex comprising:

- (i) a targeting polypeptide as defined in any one of the preceding claims; and

(ii) an antigen or a nucleic acid encoding an antigen, wherein the antigen or encoded antigen is selected from a pathogenic antigen, auto-antigen, an allergen and a cancer antigen.

5 The invention also provides a virus comprising a targeting polypeptide of the invention

In addition, the invention provides:

- a nucleic acid molecule comprising a polynucleotide sequence encoding a targeting polypeptide and an antigen selected from a pathogenic antigen, auto-antigen, an allergen and a cancer antigen;
- 10 - a vector comprising a nucleic acid of the invention; and
- a cell comprising a nucleic acid or a vector of the invention or infected with a virus of the invention.

The invention also provides a method of loading antigen presenting cells comprising contacting an antigen presenting cell with a complex or virus of the  
15 invention. An antigen presenting cell which has been loaded with a complex or virus is also provided.

The invention additionally provides:

- a pharmaceutical composition comprising a complex of the invention, a nucleic acid encoding a targeting polypeptide and antigen of a complex of the  
20 invention, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus of the invention or an antigen presenting cell of the invention and a pharmaceutically acceptable carrier or diluent;
- a vaccine comprising a complex of the invention, a nucleic acid encoding the targeting polypeptide and antigen of a complex of the invention, a vector comprising  
25 such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus of the invention or an antigen presenting cell of the invention; and
- a complex of the invention, a nucleic acid encoding a targeting polypeptide and antigen of a complex of the invention, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus of the invention, or an antigen  
30 presenting cell of the invention for use in a method of treatment of the human or animal body by therapy.

The invention also provides for the use of a nucleic acid encoding a targeting polypeptide and antigen of a complex of the invention, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus of the invention

or an antigen presenting cell of the invention in the manufacture of a medicament for use in immunisation.

The invention further provides a method of immunising a subject, the method comprising administering an effective amount of a complex of the invention, a nucleic acid encoding a targeting polypeptide and antigen of a complex of the invention, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus of the invention or an antigen presenting cell of the invention to a subject.

The invention also provides an agent for immunising a subject, the agent comprising a complex of the invention, a nucleic acid encoding a targeting polypeptide and antigen of a complex of the invention, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus of the invention or an antigen presenting cell of the invention.

#### **Brief Description of the Figures**

Figure 1 - Panel (a) shows the structure of SSL7, shaded from white (N-terminal) to dark (C-terminal). Panel (b) shows the structure of SSL7 (dark) optimally superposed on that of SSL5 (grey) and the same structure is shown on the right rotated by 90°. Panel (c) shows the structure of SSL7 (dark) optimally superposed on that of SPEC (grey) and the same structure is shown on the right rotated by 90°. Panel (d) shows the SSL7 dimer.

Figure 2 - shows residues in the SSL7 dimer interface with residue numbers being given on the x-axis, and buried surface area ( $\text{\AA}^2$ ) on the y-axis. The two forms are shown by broken and unbroken lines.

Figure 3 - Panel (a) shows the results of FACS analysis of PBMCs (peripheral blood mononuclear cells) incubated with varying concentrations of SSL-7-FITC at 4°C and 37°C. The graph is representative of ten separate experiments carried out with PBMCs from healthy donors. Panel (b) shows the results of competition experiments between SSL9-FITC (top) and SSL7-FITC (bottom) with various molecules for binding to PBMCs. Results shown are from one representative experiment of three.

Figure 4 - shows the results of FACS analysis of PBMC stained with SSL7 (top) or SSL9 (bottom) and various markers. The results shown are representative of five separate experiments.

Figure 5 – shows the results of FACS analysis of PBMC-derived dendritic cells stained with SSL9, SSL7 or no stain. The results are representative of a set of three separate experiments on cells from healthy donors.

Figure 6 – shows that SSL and SSL9 interact selectively with dendritic cells. FACS results for unpurified Dendritic cells incubated with SSL7-FITC or SSL9-FITC and then stained for CD1a are shown. The numbers show the percentage of SSL positive cells that were also CD1a positive.

Figure 7 – shows that SSLs do not alter Dendritic cell morphology or cell surface phenotype. Panel (A) shows the morphology of Dendritic cells treated with SSL proteins. Results for lipopolysaccharide (LPS) and peptidoglycan (PG) are shown as positive controls. Results are representative of three experiments. Panel (B) shows FACS results for phenotypic analysis of dendritic cells treated with SSL proteins. Data are shown for expression of cell surface molecules on Dendritic cells that have been treated with SSL7 or SSL9. LPS and PG were used as positive controls. Expression of the indicated markers is shown by the solid histograms, whereas cells stained with relevant control mAb are indicated by the open line histograms. The numbers on each histogram correspond to the median fluorescence intensity (MFI) of mAb staining. Results shown are from one donor and are representative of similar data obtained from experiments carried out with dendritic cells from four different donors.

Figure 8 – shows endocytosis of FITC-Dextran by Dendritic cells exposed to SSL proteins. The results of one of three separate experiments are shown.

Figure 9 – shows that the effect of SSL protein on the T cell stimulatory capacity of dendritic cells. Panel (A) shows the effect of SSL7 (black bars) or SSL9 (empty bars) or medium alone (grey bars) on stimulation of autologous T cells in the presence or absence (M) of purified protein derivative (PPD). Data shown are mean  $\pm$  SEM of 3 experiments.

Figure 10 – shows autologous T cell responses to Dendritic cells loaded with SSL7 or SSL9 protein. Data are mean  $\pm$  SD of triplicate cultures from individual experiments.

Figure 11 – shows the effects of SSL protein on cytokine production. Purified protein derivative is used as control.

Figure 12 – shows antibody responses in human sera. Panel (A) show levels of SSL7 or SSL9 measured by ELISA. Results for a polyclonal rabbit antibody raised

against purified His-tagged SSL7 are included as a positive control. Data are representative of 3 experiments. Panels (B) and (C) show the results of a competitive ELISA with plates coated with either SSL7 (B) or SSL9 (C) where serum diluted 1:2000 (final dilution) was mixed with varying concentrations of SSL7, SSL9 or  
5   Emblp32 as indicated.

### **Brief Description of the Sequences**

SEQ ID No: 1 provides the nucleotide sequence of a genomic fragment comprising the pathogenicity island SaPI<sub>n</sub>2 from *S. aureus* strain N315 which  
10   includes the *ssl1* to *ssl5* and *ssl7* to *ssl11* genes.

SEQ ID Nos: 2 to 11 provide the amino acid sequences of SSL1 to SSL7 and SSL7 to SSL 11 from *S. aureus* strain N315 and also indicate the location of the CDS of the genes in SEQ ID No:1.

SEQ ID No: 12 provides the nucleotide sequence of a genomic fragment from  
15   *S. aureus* strain N315, which includes the *ssl12* to *ssl14* genes. The sequence indicated is the complement of the coding strand. The order of the genes in the indicated sequence going from 5' to 3' is *ssl14*, *ssl 13* and *ssl12*. The complementary sequence is given by SEQ ID No:103.

SEQ ID Nos: 13 to 15 provide the amino acid sequences of SSL14, SSL 13  
20   and SSL12 respectively from *S. aureus* strain N315 and also indicate the location of the CDS of the genes in SEQ ID No:12. The start of each coding sequence indicated is the higher nucleotide position listed.

SEQ ID No: 16 provides the nucleotide sequence of a genomic fragment comprising the pathogenicity island SaPI<sub>n</sub>2 from *S. aureus* strain Mu50, which  
25   includes the *ssl1* to 3, *ssl5* and *ssl7* to *ssl11* genes.

SEQ ID Nos: 17 to 25 provide the amino acid sequences of SSL1 to SSL3, SSL5 and SSL 7 to 11 respectively from *S. aureus* strain Mu50 and also indicate the location of the CDS of the genes in SEQ ID No: 16.

SEQ ID Nos: 26 to 36 provide the amino acid sequences of SSL1 to SSL11  
30   respectively from *S. aureus* strain MW2 and also indicate the location of the CDS in SEQ ID No: 103.

SEQ ID No: 37 provides the nucleotide sequence of a genomic fragment from *S. aureus* strain NCTC8325, and includes the sequences of the *ssl1* to *ssl11* genes. The sequence indicated is the complement of the coding strand. The order of the

genes in the indicated sequence going from 5' to 3' is *ssl11 to ssl1*. The complementary DNA sequence is given by SEQ ID No: 104.

SEQ ID Nos: 38 to 48 provide the amino acid sequences of SSL1 to 11 respectively from *S. aureus* strain NCTC8325 and also indicate the location of the CDS of the genes in SEQ ID No:37. The start of each coding sequence indicated is the higher nucleotide position listed.

SEQ ID No: 49 provides the nucleotide sequence of a genomic fragment from *S. aureus* strain NCTC8325, and includes the sequences of the *ssl12 to ssl14* genes.

SEQ ID Nos: 50 to 52 provide the amino acid sequences of SSL12 to 14 respectively from *S. aureus* strain NCTC8325 and also indicate the location of the CDS of the genes in SEQ ID No:49.

SEQ ID No: 53 provides the nucleotide sequence of a genomic fragment from *S. aureus* strain EMRSA 16(252), which includes the sequences of the *ssl1 to ssl5, ssl7, ssl9 to ssl11* genes. The sequence indicated is the complement of the coding strand. The order of the genes in the indicated sequence going from 5' to 3' is *ssl11 to ssl9, ssl7 and ssl5 to ssl1*. The complementary DNA sequence is given by SEQ ID No:105.

SEQ ID Nos: 54 to 62 provide the amino acid sequences of SSL1 to SSL5, SSL7, SSL9 to SSL11 respectively from *S. aureus* strain EMRSA 16(252) and also indicate the location of the CDS of the genes in SEQ ID No:53. The start of each coding sequence indicated is the higher nucleotide position listed.

SEQ ID No: 63 provides the nucleotide sequence of a genomic fragment from *S. aureus* strain EMRSA 16(252) and includes the sequences of the *ssl12 to 14* genes. The sequence indicated is the complement of the coding strand. The order of the genes in the indicated sequence going from 5' to 3' is *ssl14 to ssl12*. The complementary sequence is given by SEQ ID No:106.

SEQ ID No: 64 to 66 provide the amino acid sequences of SSL12 to 14 respectively from *S. aureus* strain EMRSA 16(252) and also indicate the location of the CDS of the genes in SEQ ID No:63. The start of each coding sequence indicated is the higher nucleotide position listed.

SEQ ID No: 67 provides the nucleotide sequence of a genomic fragment from *S. aureus* strain MSSA-476 which includes the *ssl11 to ssl11* genes.

SEQ ID No: 68 to 78 provide amino acid sequences of SSL1 to SSL11 respectively from *S. aureus* strain MSSA-476 and also indicate the location of the

CDS of the genes in SEQ ID No:67.

SEQ ID No: 79 provides the nucleotide sequence of a genomic fragment from *S. aureus* strain MSSA-476 which includes the sequences of the *ssl12* to *ssl14* genes.

5 SEQ ID No: 80 to 82 provides the amino acid sequences of SSL12 to SSL 14 from *S. aureus* strain MSSA-476 and also indicate the location of the CDS of the genes in SEQ ID No:79.

SEQ ID Nos: 83 and 84 provide the nucleotide and amino acid sequences respectively of *ssl11* from *S. aureus* strain COL.

10 SEQ ID Nos: 85 and 86 provide the nucleotide and amino acid sequences respectively of *ssl12* from *S. aureus* strain COL.

SEQ ID Nos: 87 and 88 provide the nucleotide and amino acid sequences respectively of *ssl13* from *S. aureus* strain COL.

SEQ ID Nos: 89 and 90 provide the nucleotide and amino acid sequences respectively of *ssl14* from *S. aureus* strain COL.

15 SEQ ID Nos: 91 and 92 provide the nucleotide and amino acid sequences respectively of *ssl9* from *S. aureus* strain COL.

SEQ ID Nos: 93 and 94 provide the nucleotide and amino acid sequences respectively of *ssl10* from *S. aureus* strain COL.

20 SEQ ID Nos: 95 and 96 provide the nucleotide and amino acid sequences respectively of *ssl11* from *S. aureus* strain COL.

SEQ ID Nos: 97 and 98 provide the nucleotide and amino acid sequences respectively of *ssl12* from *S. aureus* strain COL.

SEQ ID Nos: 99 and 100 provide the nucleotide and amino acid sequences respectively of *ssl13* from *S. aureus* strain COL.

25 SEQ ID Nos: 101 and 102 provide the nucleotide and amino acid sequences respectively of *ssl14* from *S. aureus* strain COL.

SEQ ID No: 103 provides the complementary sequence to SEQ ID No: 12.

SEQ ID No: 104 provides the complementary sequence to SEQ ID No: 37.

SEQ ID No: 105 provides the complementary sequence to SEQ ID No: 53.

30 SEQ ID No: 106 provides the complementary sequence to SEQ ID No: 63.

SEQ ID No: 107 provides the nucleotide sequence of a genomic fragment from *S.aureus* strain MW2 which includes the *ssl1* to *ssl11* genes. The amino acid sequences of the encoded proteins are provided by SEQ ID Nos: 26 to 36 which indicate the position of the coding sequences in SEQ ID NO: 107.

**Detailed Description of the invention**

Throughout the present specification and the accompanying claims the words "comprise" and "include" and variations such as "comprises", "comprising",  
5 "includes" and "including" are to be interpreted inclusively. That is, these words are intended to convey the possible inclusion of other elements or integers not specifically recited, where the context allows. In some cases, where specific constituents are recited, the embodiment may, for example, consist essentially of such constituents.

The present invention is based on the finding that SSLs are able to target  
10 themselves to antigen presenting cells (APCs). The targeting to antigen presenting cells leads to the SSLs entering the antigen presenting pathway of the antigen presenting cells and being presented on the surface of the antigen presenting cells. This enhances the immunogenicity of the SSLs. The invention uses targeting polypeptides employing, or based on, the sequence of SSLs to target a chosen antigen  
15 to antigen presenting cells to enhance the immunogenicity of the chosen antigen. In effect, the targeting polypeptides are used to deliver a chosen antigen to an antigen presenting cell.

In some instances the targeting polypeptides are used to deliver a nucleic acid molecule encoding the antigen, rather than the antigen itself. The nucleic acid  
20 molecule will then give rise to expression of the antigen in the antigen presenting cell and the subsequent presentation of the antigen.

The invention employs targeting polypeptides comprising the polypeptide sequence of an SSL, a fragment of an SSL, or a variant sequence based on either, to target a chosen antigen or encoding nucleic acid to an APC. The invention uses  
25 complexes comprising the targeting polypeptide and antigen or an encoding nucleic acid to achieve delivery of the chosen antigen to an antigen presenting cell.

***Targeting polypeptides***

The targeting polypeptides employed comprise a staphylococcal superantigen-like protein (SSL), a fragment thereof or a variant of either, where the SSL, fragment  
30 or variant has the ability to target the targeting polypeptide to an antigen presenting cell (APC). The ability to target a polypeptide to an antigen presenting cell is referred to herein as targeting activity.

The targeting polypeptide may comprise the sequence of any naturally occurring SSL polypeptide. Such polypeptides may typically be isolated from *Staphylococcus aureus*.

- 5 The amino acid sequences of SSLs 1 to 14 from a variety of strains of *Staphylococcus aureus* are provided herein and any of these sequences may be present in a targeting polypeptide of the invention. The Table below indicates the corresponding SEQ ID Nos for the SSLs.

**Table – SSLs**

<b><i>Staphylococcus aureus</i> strain and SEQ ID No for SSL</b>							
SSL	N315	Mu50	MW2	NCTC 8325	EMRSA 16(252)	MSSA 476	COL
1	2	17	26	38	54	68	84
2	3	18	27	39	55	69	86
3	4	19	28	40	56	70	88
4	5		29	41	57	71	90
5	6	20	30	42	58	72	
6			31	43		73	
7	7	21	32	44	59	74	
8	8	22	33	45		75	
9	9	23	34	46	60	76	92
10	10	24	35	47	61	77	94
<b><i>Staphylococcus aureus</i> strain and SEQ ID No for SSL</b>							
SSL	N315	Mu50	MW2	NCTC 8325	EMRSA 16(252)	MSSA 476	COL
11	11	25	36	48	62	78	96
12	15			50	64	80	98
13	14			51	65	81	100
14	13			52	66	82	102

The SSL sequence employed may be one or more of SSL1 to SSL 14. In a preferred case where, for example, it is desired to use the sequence of an SSL9, the SSL9 will be have the sequence of one of the SSL9 indicated in Table 1 i.e. will be selected from SEQ ID Nos: 9, 23, 34, 46, 60, 76 and 92. In a preferred instance where is desired to employ the sequence of a SSL5, the SSL5 may be selected from the sequences of SEQ ID Nos: 6, 20, 30, 42, 58, and 72. In another preferred case, where it is desired to employ the sequence of a SSL 7, the SSL7 may be selected from the sequences of SEQ ID No: 7, 21, 32, 42, 59 and 74. Thus Table 1 indicates examples of preferred sequences for a particular SSL.

In a particularly preferred embodiment the SSL sequence employed may be SSL5, SSL7 and/or SSL9 and may preferably be one of the specific SSL 7 and SSL9 molecules whose sequence is provided herein. The SSL sequence may be a fragment of such a sequence. The SSL sequence employed may be a variant of the specific SSL 7 and SSL9 molecules whose sequence is provided herein or may be a variant of a fragment of such a sequence.

In some cases, the SSL employed may be an allelic form of an SSL as SSL genes from different *Staphylococcus aureus* strains vary in sequence. In some instances the SSL may be from one of the *Staphylococcus aureus* strains MW2, NCTC6571, FRI326 or NCTC8325, N315, Mu50, EMRSA 16(252), MSSA-476 or COL. In a particularly preferred embodiment the SSL may be from NCTC6571.

The targeting polypeptide may comprise a fragment of a naturally occurring SSL where the fragment retains the ability to target the targeting polypeptide to an antigen presenting cell. Such fragments may comprise subregions of any one or more of the sequences of any of SSL1 to SSL14 and in particular of the sequences provided herein of particular SSL1 to SSL14 molecules.

The targeting polypeptide may, for example, contain a sub-region of an SSL that is from 25 to 200, preferably from 35 to 175, even more preferably from 50 to 150 and even more preferably 75 to 125 amino acids in length. The targeting polypeptide may comprise a sub-region of an SSL that is 220 or less, preferably 180 or less, more preferably 160 or less, still more preferably 140 or less and even more preferably 120 or less amino acids in length. Such fragments may be derived from any SSL and in particular from the amino acid sequences indicated herein for SSL1 to SSL14. In some cases, the subregion may comprise at least 30, preferably at least 50,

more preferably at least 75 and even more preferably at least 100 amino acids from the SSL. In a preferred case the fragment may be from SSL5, SSL7 and/or SSL9. In a particularly preferred case the fragment may be from SSL9 and/or SSL7.

5 The targeting polypeptide may comprise a variant sequence of an SSL or a fragment thereof. Such variant sequences will be able to target themselves to antigen presenting cells. Any suitable variant polypeptide capable of directing the targeting polypeptide to an antigen presenting cell may be employed. In some cases, the variant will have at least 25%, preferably at least 30%, more preferably at least 35%, still more preferably at least 40% and even more preferably at least 45% amino acid  
10 sequence identity to an SSL and in particular to an amino acid sequence of any one or more of the specific sequences provided herein for SSL 1 to 14. Thus the variant may have the specified level of sequence identity with the equivalent SSL whose sequence is provided herein.

The level of amino acid sequence identity may be at least 50%, more  
15 preferably at least 55%, even more preferably at least 60% and still more preferably at least 65% amino acid sequence identity. The variant may, for example, have at least 75%, preferably at least 80%, more preferably at least 85%, still more preferably at least 90% sequence identity and even more preferably at least 95% sequence identity to an SSL and in particular to one or more of the specific SSL1 to 14 molecules  
20 whose sequence is provided herein.

In some cases the SSL employed may be an allelic variant of a known SSL gene including an allele of any of SSLs 1 to 14. Such variants may have a high degree of sequence identity to a known SSL allele and particular to one of those provided herein. Thus in some cases the variant may have at least 85%, preferably at least  
25 95%, more preferably at least 97%, and even more preferably at least 99% sequence identity to any of the SSLs mentioned herein. The amino acid sequences of alleles of SSL5, SSL7 and SSL9 are indicated in Table 1 and the targeting polypeptide may comprise one or more of these sequences.

In instances where the targeting polypeptide comprises a variant sequence, the  
30 variant sequence may have one of the levels of sequence identity specified herein to more than one SSL. Thus, for example, the invention encompasses a variant of an SSL or a variant of a fragment of an SSL that has one of the specified levels of sequence identity to all of the SSLs whose sequence is provided herein or to a particular SSL. The sequence may have one of the specified levels sequence identity

to all of the alleles provided herein for a particular SSL whose sequence is provided herein.

In some cases a variant sequence may have one of the specified levels of sequence identity to at least two, preferably at least three, more preferably at least five and even more preferably at least six of the sequences provided herein for a particular SSL. In a preferred embodiment it will have one of the levels of sequence identity specified to all of the sequences provided herein for SSL5, SSL7 and/or SSL9. The variant may in particular have one of the specified levels of sequence identity to all of the sequences provided herein for SSL7 and/or SSL9. In a preferred case, the variant may have the specified level of sequence identity to the alleles from strain NCTC6571.1

The targeting polypeptide may comprise a variant sequence of a fragment of an SSL. Thus the fragment may be any of the lengths referred to herein. It may have any of the degrees of sequence identity referred to herein. In general a variant sequence may have the specified level of sequence identity over the whole of the variant, over at least 20, preferably at least 50, more preferably at least 75, and even more preferably over at least 100 amino acids. In the case of fragments, the sequence identity may be over any of the lengths specified herein and in particular over the entire fragment or targeting sequence. The variant or fragment has to be able to target the polypeptide to an antigen presentation cell. In cases where the targeting polypeptide comprises sequences, other than those responsible for targeting, the level of sequence identity may be over the region corresponding to the SSL, fragment thereof or variant of either. The specified level of sequence identity may be over the minimum region in the polypeptide necessary for targeting to the antigen presenting cell.

The variant may have amino acid substitutions in comparison to the normal sequence of the SSL. For example, it may have from 1, 2, 3 or more substitutions such as from 5 to 10, 10 to 20, 20 to 30 or more amino acid substitutions. The variant may have, in addition or alternatively, such numbers of amino acids deleted or inserted into it in comparison to the normal SSL. The amino acid changes may be conservative substitutions, for example according to the following Table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other.

ALIPHATIC	Non-polar	G A P
		I L V
	Polar-uncharged	C S T M
		N Q
	Polar-charged	D E
		K R
AROMATIC		H F W Y

The targeting polypeptide may comprise additional sequences to those  
 5 responsible for targeting to an antigen presenting cell. In some cases the targeting  
 polypeptide will consist essentially of the targeting sequences. In cases where there  
 are additional sequences present, they may serve a variety of roles. In a particularly  
 preferred embodiment, the targeting polypeptide will also comprise the antigen it is  
 desired to target to the immune cells. The antigen may be any of those discussed  
 10 herein.

Polypeptide sequences may be present which separate the various regions of  
 the targeting polypeptide. For example, regions may be present to separate the region  
 responsible for targeting from an antigen. Polypeptide sequences may be present  
 which allow for purification of the polypeptide such as, for example, a histidine tag or  
 15 an antibody recognition site. The targeting polypeptide may include an enzymatic  
 cleavage site and in particular a protease recognition site may be present in the  
 targeting polypeptide. It may be desirable to be able to remove sequences from the  
 polypeptide such as those used to purify the polypeptide and cleavage sites may be  
 employed.

The targeting polypeptide may lack specific sequences, for example because they have been cleaved after they are no longer of use. Thus in a preferred instance the targeting polypeptide may, for example, lack a Histidine tag and hence lack four, five, six, seven, eight, ten or more consecutive histidine residues typically at the terminus of the protein. The targeting polypeptide may lack an antibody recognition sequence used in purification, a reporter sequence and/or sequences such as thioredoxin or Pho A sequences. In some embodiments of the invention such sequences will not be present at any stage and will not be encoded by the nucleic acids of the invention.

The targeting polypeptide may comprise viral sequences. A targeting polypeptide of the invention may be used to target a viral particle to an antigen presenting cell. In such cases, the targeting polypeptide will typically be provided wholly or partially on the surface of the virus to allow the targeting sequence to target the virus to an antigen presenting cell. The targeting polypeptide may increase the affinity of the viral particle for the antigen presenting cell and may mean that the viral particle is more selective in binding to antigen presenting cells as opposed to other cell types.

The viral sequences may be, or comprise sequences from, surface proteins or polypeptide sequences which allow the targeting polypeptide to be displayed on the surface of the virus. The targeting polypeptide may comprise trans-membrane sequences to allow the targeting polypeptide to be present in a membrane and in particular in a viral membrane.

The targeting sequence may comprise linker sequences to allow different domains in the targeting polypeptide to function. In one case the targeting polypeptide may comprise a proline rich linker and in particular all or part of the proline rich sequence from the 4070A protein from a leukaemia virus surface protein. In cases where the targeting polypeptide is employed to target a viral particle to an antigen presenting cell the targeting sequences may be present in the targeting polypeptide together with the sequences necessary for fusion of the membranes of the virus and antigen presenting cell to allow entry of the virus into the cell.

The targeting polypeptide may comprise a protease cleavage site which will allow cleavage of the part of the targeting polypeptide and in particular a protease site which may be cleaved by a protease present on the surface of the antigen presenting cell.

In a preferred embodiment, the targeting polypeptide may comprise the sequence of SSL 5, 7, or 9 and in a particularly preferred case the sequence of SSL 7 or 9 may be employed. In some cases the sequence of SSL 7 may be employed, particularly where it is desired to target B cells such as CD19<sup>+</sup> B cells. In other cases, such as for example where it is not desired to target B cells, other SSLs may be employed and in particular SSL9 may be employed. In each case fragments of such sequence or variants of either may be employed including any of those referred to herein.

In a preferred embodiment of the invention the targeting polypeptide comprises:

- (a) a SSL polypeptide having the amino acid sequence of any of SEQ ID Nos: 2 to 11, 13 to 15, 17 to 36, 38 to 48, 50 to 52, 54 to 66, 68 to 78, 80 to 82, 84, 86, 88, 90, 92, 94, 96, 98, 100 and 102;
- (b) a fragment of any of the sequences of (a), the fragment having the ability to target the complex to an antigen presenting cell; and/or
- (c) a variant polypeptide having at least 30% amino acid sequence identity to any of the polypeptides of (a) or (b) and the ability to target the complex to an antigen presenting cell.

In another the targeting polypeptide comprises:

- (a) the sequence of SEQ ID No: 7, 21, 32, 44, 59 and/or 74;
- (b) a fragment of any of the sequences of (a), the fragment having the ability to target the complex to an antigen presenting cell; and/or
- (c) a variant polypeptide having at least 70 % amino acid sequence identity to any of the polypeptides of (a) or (b) and the ability to target the complex to an antigen presenting cell.

In another the targeting polypeptide comprises:

- (a) the sequence of SEQ ID No: 9, 23, 34, 46, 60, 76 and/or 92;
- (b) a fragment of any of the sequences of (a), the fragment having the ability to target the complex to an antigen presenting cell; and/or
- (c) a variant polypeptide having at least 70 % amino acid sequence identity to any of the polypeptides of (a) or (b) and the ability to target the complex to an antigen presenting cell.

In some cases a targeting polypeptide may comprise a plurality of sequences which individually would be able to lead to targeting to an antigen presenting cell. A minimal sequence necessary for targeting to an antigen presenting cell may be referred to as a targeting sequence. Such targeting sequences may be any of those  
5 discussed herein which have the ability to target a polypeptide to an antigen presenting cell. Thus a targeting polypeptide may comprise, for example, 1, 2, 3, 5 or more such targeting sequences. The targeting polypeptide may comprise any pair of those targeting sequences referred to herein. In some cases the targeting polypeptide may comprise different targeting sequences with different properties.

10 In some cases the targeting polypeptide will be employed to target a viral particle to an antigen presenting cell. Thus the targeting polypeptide collectively present on the virus will be able to increase the affinity and/or specificity of the viral particle for an antigen presenting cell in comparison to an equivalent virus lacking the targeting polypeptide.

### 15 *Antigens*

The targeting polypeptides of the invention are used to deliver a chosen antigen to an antigen presenting cell. The antigen may be any suitable antigen and  
20 typically will be a peptide or a polypeptide antigen. The antigen may, for example, be an antigen selected from a pathogenic antigen, an auto-antigen, an allergen and a cancer antigen

In some cases the invention may be used to deliver a nucleic acid molecule encoding an antigen to an antigen presenting cell. The nucleic acid may then be  
25 expressed in the antigen presenting cell to give rise to the antigen and presentation of the antigen. Thus the invention also encompasses the targeting of a nucleic acid molecule encoding any of the antigens mentioned herein to an antigen presenting cell.

In a preferred instance the antigen, may be an antigen from an infectious organism. The antigen may, for example, be derived from a virus, bacterium, parasite,  
30 protozoan, fungus, or prion. The antigen may be a surface antigen expressed on the surface of the pathogen or may be an intracellular antigen. The antigen may be from an intracellular pathogen or alternatively an extracellular one.

The antigen may, for example, be from a bacterium. It may be from a gram positive or a gram negative bacterium. The antigen may, for example, be: an antigen

from *Mycobacterium* (for example from *Mycobacterium leprae*, *Mycobacterium tuberculosis*, *Mycobacterium avium*, *Mycobacterium intracellulare*, *Mycobacterium kansasii*, or *Mycobacterium gordonae*); *Pseudomonas*; *Yersinia*; *Salmonella* (for example from *Salmonella typhimurium*); *Helicobacter* (for example from  
 5 *Helicobacter pylori*); *Borelia* (for example from *Borelia burgdorferi*); *Bordetella* (for example from *Bordetella pertussis* or *Bordetella parapertussis*); *Legionella* (for example from *Legionella pneumophila*); *Staphylococcus* (for example from *Staphylococcus aureus*); *Neisseria* (for example from *Neisseria gonorrhoeae* or *Neisseria meningitidis*); *Listeria* (for example from *Listeria monocytogenes*); or  
 10 *Streptococcus* (for example from *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus viridans*, *Streptococcus faecalis*, *Streptococcus bovis*, or *Streptococcus pneumoniae*).

In some cases the antigen may be from: a *Campylobacter*; *Enterococcus*; *Haemophilus* (for example from *Haemophilus influenzae*); *Bacillus* (for example from  
 15 *Bacillus anthracis*); *Corynebacterium* (for example from *Corynebacterium diphtheriae*); *Erysipelothrix* (for example from *Erysipelothrix rhusiopathiae*); *Clostridium* (for example from *Clostridium perfringens*, or *Clostridium tetani*); *Vibrio* (for example *Vibrio cholerae*); *Enterobacter* (for example from *Enterobacter aerogenes*); *Klebsiella* (for example from *Klebsiella pneumoniae*); *Pasturella* (for  
 20 example from *Pasturella multocida*); *Bacteroides*; *Fusobacterium* (for example from *Fusobacterium nucleatum*); *Streptobacillus* (for example from *Streptobacillus moniliformis*); *Shigella*; *Escherichia* (for example from *Escherichia coli*); *Rickettsia*; *Treponema* (for example from *Treponema palladium*); *Lactococcus*; *Lactobacillus*; *Brucella*; *Aeromonas*; *Francisella*; *Citrobacter*; *Rhodococcus*; *Leishmania*; or  
 25 *Strongylus* (for example from *Strongylus vulgaris*).

Examples of preferred bacterial antigens include: the *Shigella sonnei* form 1 antigen; the F1 antigen of *Yersinia pestis*; antigens from *Neisseria meningitidis* and in particular those encoded by the GNA33, GNA2001, GNA1220 and GNA1946 genes; the O-antigen of *V. cholerae* Inaba strain 569B; protective antigens of  
 30 enterotoxigenic *E. coli*, such as fimbrial antigens including colonisation factor antigens, in particular CFA/I, CFA/II, and CFA/IV and the nontoxic B-subunit of the heat-labile toxin; pertactin of *Bordetella pertussis*, adenylate cyclase-hemolysin of *B. pertussis*; fragment C of tetanus toxin of *Clostridium tetani* and the LT (heat labile enterotoxin) and ST (heat stable toxin) antigens.

In some instances the antigen may be a viral antigen. The antigen may, for example, be a viral coat protein, glycoprotein or other proteins expressed on the surface of a virus. The antigen may be from a Picornaviridae (for example from a polio virus, a hepatitis virus, an enterovirus, a coxsackie virus, a rhinovirus, or an echovirus); Caliciviridae; Togaviridae (for example from a equine encephalitis virus or a rubella virus); Flaviridae (for example from a dengue virus, an encephalitis virus, or a yellow fever virus); Coronaviridae (for example from a coronavirus); Rhabdoviridae (for example from a vesicular stomatitis virus, or a rabies virus); Filoviridae (for example from an ebola virus); Paramyxoviridae (for example from a parainfluenza virus, mumps virus, measles virus, or a respiratory syncytial virus); Orthomyxoviridae (for example from an influenza virus such as influenza types A, B and C); Bungaviridae (for example from a Hanta virus, bunga virus, phlebovirus or a Nairo virus); Arena viridae (for example from a hemorrhagic fever virus); Reoviridae (for example a rotavirus); Birnaviridae; Hepadnaviridae (for example a Hepatitis B virus); Parvoviridae; Papovaviridae (for example from a papilloma virus, or polyoma virus); Adenoviridae; Herpesviridae (for example from herpes simplex virus (HSV) 1 or 2, varicella zoster virus, cytomegalovirus or a herpes viruses); Poxviridae (for example from a variola virus, vaccinia virus, or a pox virus); or an Iridoviridae (for example from African swine fever virus).

In a preferred case the antigen may be from a Retroviridae (e. g., HTLV-I; HTLV-11; or HIV-1 (also known as HTLV-111, LAV, ARV, hTLR, etc.)). In a particularly preferred case the antigen may be one derived from HIV and in particular the isolates HIVIIIb, HIVSF2, HTVLAV, HIVLAI, HIVMN; HIV-1CM235, HIV-1; or HIV-2. In a particularly preferred embodiment, the antigen may be a human immunodeficiency virus (HIV) antigen. Examples of preferred HIV antigens include, for example, gp120, gp160 gp41, gag antigens such as p24gag and p55gag, as well as proteins derived from the pol, env, tat, vif, rev, nef, vpr, vpu or LTR regions of HIV. In a particularly preferred case the antigen may be HIV gp120 or a portion of HIV gp120. The antigen may be from an immunodeficiency virus, and may, for example, be from SIV or a feline immunodeficiency virus.

In a preferred case the viral antigen may be one from a hepatitis virus such as an antigen from hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV) or hepatitis G virus

(HGV). In a particularly preferred embodiment the antigen may be from hepatitis B virus (HBV) and may preferably be a surface or core antigen from HBV.

In another preferred case the antigen may be from a herpesvirus family.

Particular antigens include those from herpes simplex virus (HSV) types 1 and 2, such as HSV-1 and HSV-2 glycoproteins gB, gD and gH; antigens from varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV) including CMV gB and gH; and antigens from other human herpesviruses such as HHV6 and HHV7.

The antigen may be a fungal antigen, such as one from *Candida* or *Aspergillus*. In particular, it may be from *Candida albicans* or *Aspergillus fumigatus*. The antigen may be from *Sporothrix* (e.g. from *Sporothrix schenckii*), *Histoplasma* (e.g. from *Histoplasma capsulatum*) *Cryptococcus* (e.g. from *Cryptococcus neoformans*) or *Pneumocystis* (e.g. from *Pneumocystis carinii*). The antigen may be from a parasitic pathogen and may, in particular, be from Taenia, Flukes, Roundworms, Amebiasis, Giardiasis, Cryptosporidium, Schistosoma, *Pneumocystis carinii*, Trichomoniasis and Trichinosis.

In some cases the antigen may be an antigen from a prion. In particular, the antigen may be one from the causative agent of kuru, Creutzfeldt-Jakob disease (CJD), scrapie, transmissible mink encephalopathy and chronic wasting diseases, or from a prion associated with a spongiform encephalopathy, particularly BSE. The antigen may be from the prion responsible for familial fatal insomnia.

In some cases the antigen may be from a parasitic pathogens including, for example, one from the genera *Plasmodium*, *Chtamydia*, *Trypanosome*, *Giardia*, *Boophilus*, *Babesia*, *Entamoeba*, *Eimeria*, *Leishmania*, *Schistosoma*, *Brugia*, *Fasciola*, *Dirofilaria*, *Wuchereria* and *Onchocerca*. Examples of preferred antigens from parasitic pathogens to be expressed as the heterologous antigen include the circumsporozoite antigens of *Plasmodium* species, such as the circumsporozoite antigen of *P. bergerii* or the circumsporozoite antigen of *P. falciparum*; the merozoite surface antigen of *Plasmodium* species; the galactose specific lectin of *Entamoeba histolytica*; gp63 of *Leishmania* species; paramyosin of *Brugia malayi*; the triose-phosphate isomerase of *Schistosoma mansoni*; the secreted globin-like protein of *Trichostrongylus colubriformis*; the glutathione-S-transferases of *Frasciola hepatica*, *Schistosoma bovis* and *S. japonicum*; and KLH of *Schistosoma bovis* and *S. japonicum*.

In some cases the antigen may be a cancer antigen and in particular a tumour antigen. Examples of particular cancers that the antigen may be derived include those from cancers of the lung, prostate, breast, colon, ovary, melanoma, a lymphoma and leukaemia. Examples of particular tumour antigens include MART-1, Melan-A, tyrosinase, p97, beta-HCG, GaINAc, MAGE-1, MAGE-2, MAGE-4, MAGE-12, MUC1, MUC2, MUC3, MUC4, MUC18, CEA, DDC, P1A, EpCam, melanoma antigen gp75, Hker 8, high molecular weight melanoma antigen, K19, Tyr1, Tyr2, members of the pMel 17 gene family, c-Met, PSA (prostate antigen), PSM (prostate mucin antigen), PSMA (prostate specific membrane antigen), prostate secretory protein, alpha-fetoprotein, CA125, CA19.9, TAG-72, BRCA-1 and BRCA-2 antigens.

The antigen may be an auto-antigen. In particular, the antigen may be an antigen associated with an autoimmune disease. Auto-antigens include those associated with autoimmune diseases such as multiple sclerosis, insulin-dependent type 1 diabetes mellitus, systemic lupus erythematosus (SLE) and rheumatoid arthritis. The antigen may be one associated with, Sjorgrens syndrome, myotitis, scleroderma or Raynaud's syndrome. Further examples of auto-immune disorders that the antigen may be associated with include ulcerative colitis, Crohns' disease, inflammatory bowel disorder, autoimmune liver disease, or autoimmune thyroiditis. Examples of specific autoantigens include insulin, glutamate decarboxylase 65 (GAD65), heat shock protein 60 (HSP60), myelin basic protein (MBP), myelin oligodendrocyte protein (MOG), proteolipid protein (PLP), and collagen type II. In cases where the antigen is an autoantigen the antigen will typically be administered in order to promote tolerance to the auto-antigen. Although in some cases models of the diseases may be produced using the invention to be produce an immune response.

In some cases the antigen may be an allergen. The allergenic antigen may be any suitable antigen from an antigen. For example, the allergen may be from *Ambrosia artemisiifolia*, *Ambrosia trifida*, *Artemisia vulgaris*, *Helianthus annuus*, *Mercurialis annua*, *Chenopodium album*, *Salsola kali*, *Parietaria judaica*, *Parietaria officinalis*, *Cynodon dactylon*, *Dactylis glomerata*, *Festuca pratensis*, *Holcus lanatus*, *Lolium perenne*, *Phalaris aquatica*, *Phleum pratense*, *Poa pratensis* or *Sorghum halepense*. The allergen antigen may be from a tree, such as, for example, from *Phoenix dactylifera*, *Betula verrucosa*, *Carpinus betulus*, *Castanea sativa*, *Corylus avellana*, *Quercus alba*, *Fraxinus excelsior*, *Ligustrum vulgare*, *Olea europea*, *Syringa vulgaris*, *Plantago lanceolata*, *Cryptomeria japonica*, *Cupressus arizonica*,

*Juniperus oxycedrus*, *Juniperus virginiana*, or *Juniperus sabinoides*. In some cases the antigen may be from an antigen from a mite such as, for example, from *Acarus siro*, *Blomia tropicalis*, *Dermatophagoides farinae*, *Dermatophagoides microceras*, *Dermatophagoides pteronyssinus*, *Euroglyphus maynei*, *Glycyphagus domesticus*,  
5 *Lepidoglyphus destructor* or *Tyrophagus putrescentiae*.

The allergen antigen may be from an animal such as, for example, from a domestic or agricultural animal. Examples of allergens from animals include those from cattle, horses, dogs, cats and rodents (e.g from rat, mouse, hamster, or guinea pig). In some cases the antigen may be from a food allergen and in others it may be  
10 from insect.

The antigen may be one involved in transplant rejection. The invention may be use to induce or promote tolerance to such an antigen in order to ameliorate or prevent transplant rejection.

Homologues of antigens may also be employed. Protein antigens employed  
15 may have homology and/or sequence identity with naturally occurring antigens, such as any of the antigens mentioned herein. They may have any of the levels of sequence identity or sequence changes specified herein.

The antigen may be a model antigen. The antigen may be one commonly used in experiments to assess immune responses. For example the antigen may be a  
20 lysozyme and in particular chicken egg lysozyme. The antigen may be ovalbumin and in particular chicken ovalbumin. Such model antigens may have the advantage that antigens, T cells and other reagents may be readily available for assessing antigen presentation by the targeted antigen presenting cell.

## 25 *Complexes*

The targeting polypeptides of the invention may be used to deliver a chosen antigen to antigen presenting cells (APCs). The targeting polypeptide and antigen are combined in the form of a complex. Thus a complex comprising a targeting  
30 polypeptide and an antigen is provided. In particular, the complex may comprise an antigen selected from a pathogenic antigen, auto-antigen, an allergen and a cancer antigen

In some cases a nucleic acid molecule encoding the antigen may be delivered to antigen presenting cells using the targeting polypeptides of the invention. Thus the  
35 invention also provides a complex comprising a targeting polypeptide and a nucleic

acid molecule encoding an antigen. The nucleic acid molecule will be delivered to the antigen presenting cell and result in expression of the antigen in the antigen presenting cell. Thus reference herein to an antigen includes a nucleic acid molecule encoding such an antigen.

5 A complex of the invention may comprise:

- a targeting polypeptide; and
- an antigen and/or a nucleic acid molecule encoding an antigen.

The targeting polypeptide and antigen may be joined together by any suitable means that ensures that the antigen is also targeted to the antigen presenting cell.

10 Preferably, the targeting polypeptide and antigen may be present in the same polypeptide. Thus in some cases the antigen and targeting polypeptide may be directly fused to each other in a single polypeptide. In others the two may be present in the same polypeptide, but separated by an intervening sequence. For example, they may be separated by from 1 to 50, preferably from 5 to 25, more preferably from 10 to 20  
15 amino acids.

In cases where the targeting polypeptide and fusion polypeptide are present in the same polypeptide, the targeting polypeptide may be separated by a sequence designed to be cleaved by a proteases in the antigen presenting cell in order to allow the two to be separated. In some cases where the targeting polypeptide and antigen  
20 are present in the same polypeptide, there may be a plurality of antigens in the polypeptide. For example, the same antigenic sequence may be repeated several times in the polypeptide such as from 1 to 50, preferably from 2 to 25, and more preferably from 5 to 10 times. In some cases the polypeptide may therefore comprise repeats of the same epitope or a group of epitopes from a particular antigen. In other cases  
25 different antigens may be present in the polypeptide. For example, two, three, four, five or more of any of the antigens mentioned herein may be present in the same polypeptide. The antigens may be from the same source or different source, and may, for example, be from different organisms.

The targeting polypeptide and antigen may not be part of the same  
30 polypeptide. For example, they may be joined together by other covalent means. The targeting polypeptide and antigen may be joined together by a covalent bond, such as a covalent bond between side chains, for example by disulphide bridges. The two may be joined by a linker or other bridging molecule. In some cases the targeting polypeptide and antigen may be provided or coated on a moiety and the complex

including the targeting polypeptide, antigen and moiety may be targeted to antigen presenting cells by virtue of the presence of the targeting polypeptide.

The complex may comprise a plurality of antigens and/or targeting polypeptides. For example, a plurality of antigens may be present and may be linked to a single targeting polypeptide or alternatively multiple antigens may be present in a polypeptide with a single targeting sequence. The complex, may for example comprise one, two, three, five, ten or more antigens. The complex may comprise antigens from different sources such as antigens from different organisms. The complex may comprise antigens from different strains of the same organism, such as from different strains of the same pathogen. In some cases the complex may comprise different allelic or mutant forms of the same antigen. For example, the antigens may be different forms of an antigen that display diversity that leads to strains of that pathogen with differing pathogenicity. The complex may comprise a plurality of copies of a single epitope where the epitope sequence is either present in the same polypeptide as the targeting polypeptide or is joined to it by one of the other means discussed herein.

The complex may comprise a plurality of targeting polypeptides. For example, the complex may comprise 1, 2, 3 or more targeting polypeptides, such as for example from 5 to 50, more preferably from 10 to 25 or even more preferably from 15 to 20 targeting polypeptides. The ratio of targeting polypeptides to antigen and/or epitope may be for example 1:1, 1:2, 1:5, 1:10 or 1:25, in some cases the ratio may be from 1:1 to 1:75, preferably from 1:2 to 1:50 and more preferably from 1:5 to 1:25. The ratio of targeting sequences to antigen and/or epitope may have such ratios.

In a particularly preferred embodiment of the invention the targeting polypeptides in the complex may be present in a dimeric form. This may be the case, for example, where the targeting sequence is, or is a fragment thereof or a variant of either of SSL7 and/or 7 and preferably of SSL 7. In a preferred case, where a fragment or variant is employed as well as having targeting activity it may be able to form a dimer.

The complex may be chemically modified, e.g. one or more, or indeed all, of the polypeptide types in the complex may be post-translationally modified. For example, they may be glycosylated or comprise modified amino acid residues. The polypeptides in the complexes and targeting polypeptides of the invention may

comprise amino acid analogs. In some cases, one or more peptides in the complex may have been generated synthetically.

In some cases libraries of different complexes may be generated. The libraries may comprise complexes with different antigens, for example, from different  
5 pathogens. In some cases, the library may comprise complexes with antigens from the same source, such as from the same organism including any of those mentioned herein. The libraries may be encoded by libraries of nucleic acids and/or vectors of the invention. Libraries may be generated and then screened to identify those complexes showing advantageous properties.

10 In some cases the complex may comprise a nucleic acid encoding the antigen rather than the antigen itself. In particular, the complex may comprise a nucleic acid molecule capable of expressing the antigen. Any of the sequences discussed herein for expressing polypeptides may be employed to express the antigen.

The invention therefore provides a viral particle which comprises a targeting  
15 polypeptide of the invention. In a particularly preferred instance the complex may comprise a viral particle which comprises:

- (i) a targeting polypeptide;
- (ii) a nucleic acid molecule encoding an antigen.

The targeting polypeptide will preferably be wholly or partially exposed on the  
20 surface of the viral particle to allow the virus to be targeted to an antigen presenting cell.

The complex may comprise the nucleic acid in any suitable manner to allow expression of the antigen in the antigen presenting cell. The complex may comprise a liposome with the targeting polypeptide present wholly or partially on the surface of  
25 the liposome to allow targeting to antigen presenting cells.

### *Nucleic acids*

The present invention also provides a nucleic acid molecule comprising a polynucleotide sequence encoding a targeting polypeptide and an antigen and in  
30 particular an antigen selected from a pathogenic antigen, auto-antigen, an allergen and a cancer antigen.

The targeting polypeptide and antigen may be encoded by separate open reading frames (ORFs) or alternatively the nucleic acid may comprise an open reading frame encoding both the polypeptide and the antigen. In a preferred case, the

nucleotide sequence encoding the targeting polypeptide and the antigen are present in a single open reading frame.

In a particularly preferred embodiment the nucleic acid will be able to express the targeting polypeptide and antigen in the form of a polypeptide comprising both.

5 As discussed elsewhere herein the targeting polypeptide and antigen may be directly fused or alternatively may be separated by intervening sequences which are also encoded by the nucleic acid.

The nucleic acid sequence may therefore encode any of the targeting polypeptides referred to herein. Thus the nucleic acid may comprise a sequence that  
10 encodes any of the SSLs, fragments thereof, or variants of either discussed herein. The nucleic acids may also encode any of the additional polypeptide sequences referred to herein. In a preferred case the nucleic acid molecule may comprise one or more of the polynucleotides sequences of SEQ ID Nos 1, 12, 16, 37, 49, 53, 63, 67, 79, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101 and/or 107. In particular, the nucleic acid  
15 may comprise one or more of the regions of SEQ ID Nos 1, 12, 16, 37, 49, 53, 63, 67, 79, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101 and/or 107 indicated herein to represent the CDS of a particular SSL. Thus any of the nucleic acid sequences provided herein which encode one or more of SSL1 to SSL14 may be used. The sequences used from the sequences provided herein may be restricted to the coding regions indicated or  
20 may also employ other regions from the gene and in particular the whole gene.

The nucleic acid may comprise the polynucleotide sequence of one or more of the coding sequences provided herein for SSL 1 to SSL14, a fragment of one or more of those sequence or a variant sequence of such a sequence or fragment, where the sequence encodes a targeting sequence able to target the encoded polypeptide to an  
25 antigen presenting cell.

The sequence of an SSL may be modified by nucleotide substitutions, insertions or deletions. In particular, the sequences provided herein which encode a SSL may be altered in such a way. The nucleic acid sequence may, for example, comprise from 1, 2, 5, 10 or 20 such substitutions, insertions and/or deletions as long  
30 as the encoded polypeptide has targeting activity. The variant sequence may comprise from 1 to 50, preferably from 5 to 25, more preferably from 10 to 15 amino acid insertions, deletions or substitutions as long as the encoded polypeptide displays targeting activity. Degenerate substitutions may be made and/or substitutions may be

made which would result in a conservative amino acid substitution when the modified sequence is translated, for example as shown in the Table above.

The nucleic acid may comprise a sequence that has at least 25%, preferably at least 30%, more preferably at least 35% and even more preferably at least 40% sequence identity to any one or more of SEQ ID Nos 1, 12, 16, 37, 49, 53, 63, 67, 79, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101 and/or 107. In particular, they may have such a level of sequence identity to the region encoding the SSL and/or over the whole gene.

In some cases the nucleic acid may have at least 50%, preferably at least 60%, more preferably at least 70%, even more preferably at least 80% sequence identity to such sequences. In other cases the level of sequence identity may be higher, because, for example, the sequence is a natural allelic variant or an engineered variant. Thus in some instances the sequence may have at least 85%, preferably at least 90%, more preferably at least 95%, even more preferably at least 97% and still more preferably at least 99% sequence identity to any of the sequences provided herein which encode SSL1 to SSL 14. In one case, the nucleic acid molecule will have one of the specified levels of nucleotide sequence identity to all of, three of, or two of the sequences encoding SSL5, SSL7 and/or SSL9 provided herein.

The levels of sequence identity specified may be over the entire sequence encoding the targeting polypeptide or targeting sequence. They may, for example, be over from 25 to 900 nucleotides, preferably over 50 to 700 nucleotides, more preferably over 75 to 350 nucleotides and even more preferably over 100 to 250 nucleotides. Thus in some cases the level of sequence identity specified may be over a region of at least 50, preferably at least 75, for instance at least 100, at least 150, more preferably at least 200 contiguous nucleotides or most preferably over the full length of the nucleic acid encoding the targeting sequence or polypeptide.

The nucleic acid may comprise the nucleotide sequence of one or more of the sequence SSL5, SSL7 and/or SSL9 genes whose sequence is provided herein, a fragment of any of the sequences or a variant of the preceding sequences where the encoded polypeptide displays targeting activity. In a preferred instance, the nucleic acid will comprise the nucleotide sequence of one or more of the sequences provided herein which encode SSL5, SSL7 and/or SSL9. In a preferred case the nucleic acid may comprise the nucleotide sequence of SEQ ID Nos: SSL7 and/or SSL9, a fragment of either or a variant of any such sequence, where the fragment or variant retains targeting ability. In a particularly preferred case the nucleic acid will comprise

the polynucleotide sequence of a sequence provided herein encoding SSL7 and/or SSL9.

Sequence identity and comparisons may be performed in a number of ways. For example the UWGCG Package provides the BESTFIT program which can be used to calculate homology (for example used on its default settings) (Devereux *et al* (1984) *Nucleic Acids Research* 12, p387-395). The PILEUP and BLAST algorithms can be used to calculate homology or line up sequences (typically on their default settings), for example as described in Altschul (1993) *J. Mol. Evol.* 36:290-300; Altschul *et al* (1990) *J. Mol. Biol.* 215:403-10.

Software for performing BLAST analyses is publicly available through the National Centre for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pair (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighbourhood word score threshold (Altschul *et al*, 1990). These initial neighbourhood word hits act as seeds for initiating searches to find HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extensions for the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

The BLAST algorithm performs a statistical analysis of the similarity between two sequences; see e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a sequence is considered similar to another sequence if the smallest sum probability in comparison of the first sequence to the second sequence is

less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Any combination of the above mentioned degrees of sequence identity and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher sequence identity over longer lengths) being preferred. Thus, for example, a polynucleotide which has at least 90% sequence identity over 25, preferably over 30 nucleotides forms one aspect of the invention, as does a polynucleotide which has at least 95% sequence identity over 40 nucleotides.

The nucleic acids of the invention may comprise a number of another sequences in addition to that encoding the targeting polypeptide and antigen. For example, the nucleic acid may comprise sequences involved in the expression of the polypeptide it encodes such as those discussed below on the section in vectors. The nucleic acids may comprise primer sites, restriction sites, multiple cloning sites and other sequences to facilitate manipulation. The nucleic acid may comprise enhancer sequences to facilitate gene expression. The nucleic acid may comprise sequences allowing for the secretion of the encoded polypeptide or its targeting to a particular cellular compartment. In some cases the nucleic acids may also comprise a reporter gene or sequences, in other cases they may not.

The nucleic acids of the invention may be used in the production of targeting polypeptides, antigens and/or complexes of the invention. Such production may take place *in vitro*, *in vivo* or *ex vivo*. The polynucleotides may be used in recombinant protein synthesis or indeed as therapeutic agents in their own right. Polynucleotides encoding a targeting polypeptide and not an antigen or alternatively those encoding an antigen, but not a targeting polypeptide may be used to produce targeting polypeptide and/or antigen which can then be utilised to form complexes of the invention.

### *Vectors*

The present invention provides vectors comprising a nucleic acid of the invention. Thus in one instance the invention provides a vector comprising a nucleic acid sequence that encode a targeting polypeptide of the invention and an antigen. The antigen may be one selected from any of those mentioned herein and in particular may be selected from a pathogenic antigen, auto-antigen, an allergen and a cancer antigen. The vector may comprise any of the nucleic acid sequences mentioned

herein. In a preferred instance, the vector may encode the targeting polypeptide and antigen via the same open reading frame (ORF).

The invention also provides vectors which comprise a nucleic acid which encodes the chosen antigen where the nucleic acid will be targeted to an antigen presenting cell using a targeting polypeptide of the invention. Such vectors will preferably be viral vectors.

The vector may, for example, be a cloning, expression and/or viral vector. The vector may, for instance, be a plasmid vector. The vector may be a viral vector. The vector may be a shuttle vector. The vector may comprise a selectable marker, for instance an antibiotic resistance selectable marker.

In particular, expression vectors are provided which are capable of expressing a targeting polypeptide of the invention and an antigen. In a preferred instance, the vector will express a targeting polypeptide of the invention as a fusion protein with the chosen antigen. Thus the vector may be capable of expressing a complex of the invention. Alternatively the targeting polypeptide and antigen may be produced separately and then linked to form a complex or the targeting polypeptide and antigen may be expressed as a single polypeptide and then further processing is carried out to produce a particular complex. For example, individual polypeptides may be linked together or sequences may be cleaved from the expressed polypeptides. In some cases sequences used in purification may be removed by cleavage.

Expression vectors are routinely constructed in the art of molecular biology and may for example involve the use of plasmid DNA and appropriate initiators, promoters, enhancers and other elements, such as for example polyadenylation signals which may be necessary, and which are positioned in the correct orientation, in order to allow for protein expression. Other suitable vectors would be apparent to persons skilled in the art. By way of further example in this regard we refer to Sambrook *et al.*, 1989. The expression vector may be a prokaryotic or a eukaryotic expression vector. In some cases the expression vector may be used to produce the encoded protein *in vitro*. In other cases, the expression vector will be intended to generate *in vivo* expression of the encoded protein and may be used, for example, in a method of therapy.

Once coding sequences for desired proteins have been prepared or isolated, such sequences can be cloned into any suitable vector or replicon. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate

cloning vector is a matter of choice. Ligations to other sequences may be performed using standard procedures, known in the art. The vector may be, for example, plasmid, virus or phage vectors provided with a origin of replication, optionally a promoter for the expression of the polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes.

Expression of the targeting polypeptide and/or antigen will typically be driven by a promoter. The promoter will usually be chosen on the basis of the cell the expression vector is to be used in. Thus for prokaryotic expression a prokaryotic promoter will typically be used, whilst for eukaryotic expression a eukaryotic or viral promoter will typically be employed. The promoter employed may be a viral or non-viral promoter. The promoter may be a mammalian promoter, such as a cell or tissue specific promoter or alternatively a promoter expressed in a wide range of cells. Other types of regulatory elements may also be present in the vector, for example, enhancer sequences.

Mammalian promoters, such as  $\beta$ -actin promoters, may be used. Tissue-specific promoters are especially preferred. Viral promoters may also be used, for example the Moloney murine leukaemia virus long terminal repeat (MMLV LTR), the rous sarcoma virus (RSV) LTR promoter, the SV40 promoter, the human cytomegalovirus (CMV) IE promoter, adenovirus, HSV promoters (such as the HSV IE promoters), or HPV promoters, particularly the HPV upstream regulatory region (URR). Viral promoters are readily available in the art.

In one preferred case the promoter employed may be one that is capable of being expressed in an antigen presenting cell. Thus a vector comprising a promoter capable of giving rise to expression of the antigen in an antigen presenting cell is provided. A promoter which is specific for antigen presenting cells may be employed.

Additional sequences may also be present in the open reading frame encoding the targeting polypeptide as well as optionally those encoding the antigen. In some cases, sequences directing secretion or release from the cell of the targeting polypeptide may be included. Peptide sequences allowing purification of the targeting polypeptide may be included. Preferably sequences allowing cleavage may be included and, for example, may be used to release the targeting polypeptide from the sequence used to purify the expressed polypeptide.

### Cells

The invention provides cells comprising a polynucleotide or vector of the invention. Thus the invention provides for a cell comprising a polynucleotide and/or  
5 vector encoding a targeting polypeptide and an antigen. The targeting polypeptide and/or antigen may be any of those mentioned herein. Preferably, the cell will express the targeting polypeptide and antigen and in particular will express polypeptides comprising both. The cell may therefore be able to produce complexes of the invention and in one case may secrete them.

10 The invention also provides for cells that can produce viruses of the invention which have a targeting polypeptide provided on their surface either wholly or partially to allow for targeting of the viral particle to an antigen presenting cell. The invention provides a helper cell line that is capable of expressing the targeting polypeptide in such a way that it is incorporated in a viral particle. Such cells may also comprise the  
15 nucleic acid to be incorporated into the viral particles. The invention also provides cells infected with a virus of the invention.

Cells of the invention include transient, or preferably stable higher eukaryotic cell lines, such as mammalian cells or insect cells, produced using, for example, a baculovirus expression system, lower eukaryotic cells, such as yeast or prokaryotic  
20 cells such as bacterial cells. Particular examples of cells include mammalian cells such as HEK293T, CHO, HeLa and COS cells. The cells may be human cells. The cells may be, for instance, from any of the species and/or subjects mentioned herein. Preferably the cell line selected may be one which is not only stable, but also allows for normal post-translation modifications, particularly so that the antigen or epitope is  
25 in the form it would be naturally expected to be encountered as. The cell may, for example, allow normal glycosylation.

A polypeptide of the invention may be expressed in cells of a transgenic non-human animal. A transgenic non-human animal expressing a polypeptide of the invention is included within the scope of the invention. Such an animal may for  
30 example be a rodent (e.g. a mouse or rat). Preferred polypeptides of the invention may also be expressed in *Xenopus laevis* oocytes.

The sequences encoding the targeting polypeptide of the invention and/or the antigen may be introduced into a chosen cell by any suitable technique and may be generally referred to without limitation as "transformation". For eukaryotic cells,

suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For example, the calcium phosphate precipitation method of Graham and van der Eb, *Virology* 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. For various techniques for transforming mammalian cells, see Keown *et al.*, *Methods in Enzymology*, 185:527-537 (1990) and Mansour *et al.*, *Nature* 336:348-352 (1988). Nucleic acids and vectors of the invention may be introduced into target cells both *in vitro* and *in vivo*. In particular, viral based systems may be used to introduce the nucleic acids and/or vectors of the invention into cells, particularly *in vivo*.

#### *Antigen Presenting cells (APCs)*

The invention allows a chosen antigen to be delivered to an antigen presenting cell. The targeting polypeptide present in the complexes direct the complex to an antigen presenting cell. The complex is taken up by the antigen presenting cell. This means that the antigen may then be presented by the antigen presenting cell.

The invention also provides for the delivery of a nucleic acid molecule encoding an antigen through the use of a targeting polypeptide of the invention. The targeting polypeptide will preferably be on the surface of a virus. Infection of the antigen presenting cell by the virus results in the production of antigen which may then be presented by the cell.

The antigen presenting cell may be any suitable antigen presenting cell. In particular, it may be a professional antigen processing cell and will typically express MHC molecules. The antigen processing cell will typically express MHC II molecules and the complex will allow the chosen antigen, or peptides derived from it, to be presented via MHC II molecules. In some cases the antigen, or peptides derived from it, may be presented via MHC I. The antigen may be presented by both MHC I and II, typically presentation will be via MHCII. Examples of MHC I and II molecules which may be involved in presentation include HLA-A2, HLA-B62, HLA-Bw62, HLA-B35, HLA-DRB1, HLA-DRB2, HLA-DRB3, HLA-DRB5, HLA-DRB7, HLA-A25, HLA-B8, HLA-B52, HLA-DQB1, HLA-A3, HLA-A11 or HLA-B27.

Examples of antigen presenting cells that the invention may be used to target antigens to include dendritic cells, monocytes, and/or or a B cells. In particular,

monocytes and/or dendritic cells may be targeted and in a preferred instance the antigen presenting cell is a dendritic cell. The B cells may typically be CD19<sup>+</sup> B cells. The monocytes may, for example, be CD14<sup>+</sup> cells and such cells may also be CD2<sup>lo</sup> cells.

5           In cases where the antigen presenting cell is a dendritic cell it may express cell surface markers known to be characteristic of dendritic cells. In particular the dendritic cell may express CD11c, CD209 (also known as DC sign) and/or CD13. The cell may be CD14<sup>lo</sup>. In some cases the cell may express all of CD11c, CD209 and CD13 and may also be CD14<sup>lo</sup>.

10           In a preferred instance the antigen presenting cell may be a dendritic cell. Any type of dendritic cell may be targeted. Examples of dendritic cells include a monocyte derived dendritic cell, a plasmacytoid derived dendritic cell or an intersititial dendritic cell. The dendritic cell may be a langerhans cell. The dendritic cell may be one present in an organ or tissue. The dendritic cell may be one from, or present in, a  
15           mucosal surface. The dendritic cell may be an intestinal dendritic cell such as one obtained from Peyer's patches. The dendritic cell may be one present in, or obtained, from an immune tissue such as from a secondary lymph node. The dendritic cell may be present in, or obtained from the spleen or a lymph node.

          The term antigen presenting cell covers any cell that can present an antigen  
20           targeted to it via a complex of the invention. In some cases antigen presenting cells may have different stages in their development during which they, for example, predominately take up antigen, rather than present it. For example, it is thought that dendritic cells may have immature stages characterised by the uptake of large amounts of potential antigens and more mature stages characterised by lower amounts  
25           of antigen uptake, but increased amounts of antigen presentation of the antigens they acquired earlier. The dendritic cell may, for example, be present in the periphery and effectively collect antigen and then move to areas such as secondary lymphoid tissues to present antigens. In one case the invention encompasses targeting of an antigen to an immature antigen presenting cell subsequently capable of presenting the antigen it  
30           has taken up, such as to an immature dendritic cell.

          In some cases it may be desirable to target immature dendritic cells in instances where it is desired to induce tolerance. In particular, delivery of antigen by employing a targeting polypeptide of the invention in the absence of a stimuli which induces or promotes dendritic cell maturation can result in tolerance. To achieve

tolerance the targeting polypeptide may preferably be used to deliver the chosen antigen or nucleic acid encoding the antigen in the absence of an adjuvant. An advantage of employing the targeting polypeptide of the invention to induce tolerance is that unlike many methods for inducing tolerance a large dose of antigen is not required. Steinman *et al* (2003) Ann. Rev. Immunol., 21:685-711 discusses the induction of tolerance via dendritic cells and the methods and assays discussed therein may be employed when inducing tolerance using the methods of the invention. In situations where it is desired to induce tolerance peripheral dendritic cells may preferably be targeted.

Conversely, in situations where it is desired to promote an immune response against the antigen, the targeting polypeptide may preferably be employed with an adjuvant and in particular one which induces or promotes stem cell maturation such as aluminium hydroxide.

In some cases the antigen cells may be manipulated *in vitro* and this may allow control of whether the cells are exposed to stimuli which promote dendritic cell maturation. Thus by ensuring that the cells are not exposed to stimuli responsible for inducing maturation the resultant cells may be used to induce tolerance. In other cases the cells will be exposed to stimuli which promote dendritic cell maturation and hence the cells can be used to promote an immune response when they are transferred to a subject.

Antigen presenting cells may be isolated from any suitable source including any of those mentioned herein. For example, such cells may be isolated from the white cells of the blood. Methods based on cell density such as LYMPHOPREP<sup>TM</sup> and centrifugation may be employed. The cells may be isolated using a variety of techniques including antibody based techniques. The cells may be isolated using negative and positive selection techniques based on surface markers which present and/or those that are not present on antigen presenting cells. In some cases, antigen presenting cells may be obtained by exposing other cells, such as precursor cells, to appropriate stimuli.

Dendritic cells may be obtained by treating monocytes with appropriate stimuli such as GM-CSF and/or IL-4. For example, cells may be culture in the presence of 10 to 500 ng/ml, preferably from 25 to 200 ng/ml, more preferably from 50 to 150 ng/ml and even more preferably in the presence of 100 ng/ml GM-CSF. In particular, human and preferably recombinant human GMCSF may be employed. The

cells may be cultured in the presence of 10 to 250 ng/ml, preferably from 25 to 150 ng/ml, more preferably from 40 to 70 ng/ml and even more preferably in the presence of 50 ng/ml IL-4. T and B lymphocytes may be removed by appropriate selection such as on the basis of the markers CD3, CD2 and/or CD19.

5           In other cases plasmacytoid cells may be induced to differentiate into dendritic cells by exposure to IL-3. In some instances immature antigen presenting cells may be induced to mature using appropriate stimuli. Such treatments will typically be performed *in vitro*.

10           The antigen presenting cells may be treated *ex vivo*. Thus the cells may be recovered from a subject loaded with antigen using the methods of the invention and then used therapeutically. The invention provides loaded antigen presenting cells. The invention provides antigen presenting cells which have been infected by a viral particle of the invention.

## 15    Assays

          The targeting polypeptides of the invention are utilised to target antigens to antigen presenting cells. This targeting ability and the downstream effects of targeting can be assessed in a number of ways.

20           The ability of a polypeptide to target itself, and hence a complex, to an antigen presenting cell can be assessed using any suitable technique. In one case *in vitro* assays may be performed to monitor binding of the targeting polypeptide to an antigen presenting cell. The targeting polypeptide may be labelled and the binding to the antigen presenting cell followed. The targeting polypeptide may be labelled with a fluorochrome and its binding to an antigen presenting cell assessed using techniques  
25           such as flow cytometry. Binding at 4°C and 37 °C may be compared to help demonstrate that the binding is specific.

          The ability of the polypeptide under study to target to an antigen presenting cell may be compared with a polypeptide known to have targeting ability. For example, where the test targeting polypeptide is based on the sequence of a particular  
30           SSL the ability to compete against that SSL may be compared using two or more different labels. The ability of a targeting polypeptide to compete against varied concentrations of a polypeptide with known binding activity may be measured. In this way polypeptides with binding activity may be identified and the level of targeting ability that they display quantified. Such assays will typically be performed *in vitro*

but the ability of antigen cells *in vivo* or *ex vivo* conditions to take up test polypeptides and/or complexes may also be measured. Again, labelling may be used to study delivery to antigen presenting cells.

5 The ability of a targeting polypeptide to enter the antigen presentation pathway may also be assessed. The presence of the targeting polypeptide in a subcellular compartment associated with antigen presentation may be measured. Labelling may be used to achieve this. Confocal microscopy can be used to confirm that the label, and hence the targeting polypeptide, has entered the cell. Stains for subcellular compartments associated with antigen presentation may be employed. For  
10 example, Texas red dextran staining may be employed to identify such compartments and co-staining may be used to confirm the presence of the targeting polypeptide in such compartments. In addition, the association of the complex, or part of it, in the same regions as MHC molecules, particularly MHC II may be examined.

The ability of a particular targeting polypeptide to lead to antigen presentation  
15 may be studied. Thus the presence of peptide sequences from the antigen being presented on the cell surface by MHC may be studied. In particular, presentation by MHCI may be examined. Techniques may be used to elute peptides bound to MHC and identify those originating from the chosen antigen. The degree of peptide presented when the antigen is provided on its own and when it is provided as part of a  
20 complex may be compared. The ability of particular polypeptides to lead to antigen presentation may be compared including using control polypeptides with known targeting ability. Thus the ability of two polypeptides to lead to presentation of the same antigen may be compared.

In some assays the downstream effects of antigen presentation may be  
25 measured. Thus assays may involve an antigen presentation cell being loaded using test polypeptides and then antigen presentation to a second cell may be measured. The ability of different targeting polypeptides and complexes to bring about the downstream effects of antigen presentation may be measured including controls with known activity. Such assays may be performed *in vitro* and, for example, they may be  
30 performed using an antigen presenting cell and a T cell known to have a receptor specific for an epitope present in the antigen. The T cell may be any suitable T cell. It may be a CD4<sup>+</sup> or CD8<sup>+</sup> T cell and in particular may be a CD4<sup>+</sup> T cell. The downstream effects of antigen presentation may also be measured *in vivo*

Downstream effects of antigen presentation include activation of the T cell. Various signal transduction effects associated with activation of the T cell may be measured. The activation may include the differentiation and/or proliferation of the T cell. Thus the number and proliferation of the T cells may be measured, using, for example, suitable labelling techniques or by measuring cell number. The expansion of particular subsets of T cells may be measured. For example, by flow cytometry. Assays involving autologous and/or allogenic T cells may, for instance be employed, including any of those mentioned herein.

The activation of the T cells may lead to the release of cytokines. For example, the activation may lead to the release of interleukins (e.g. IL-2, IL-4, IL-5, IL-6 and/or IL-10), IFN $\gamma$  and/or TNF- $\beta$ . In some cases antigen presentation may lead to a Th1 type response. Thus the cytokines released may be, or predominately be, IL-2, IFN $\gamma$  and/or TNF- $\beta$ . In other cases the response may be a Th2 type response. The cytokines released may be, or predominately be, IL-3, IL-4, IL-5, IL-6 and/or IL-10. The T cells may release factors that stimulate their own proliferation such as IL-2 and/or IL-4. Particular complexes and substances of the invention may be designed to give particular responses such as any of those mentioned herein. The release of such factors can be studied both *in vitro* and *in vivo* using techniques such as ELISA to measure levels of such compounds. In some cases, the cytokine levels measured may be one or more, or indeed all, of IFN $\gamma$ , IL-10 and IL-13.

The presence and/or number of effector T cells and memory T cells may be assessed. In addition various downstream effects may be measured. The number and activity of CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells may be measured and in particular those specific for the chosen antigen. Antibody responses may be assessed in terms of the amount and types of antibody produced. In one preferred instance, the delivery of the chosen antigen using the targeting sequences of the invention leads to an antibody response against the chosen antigen. Any of the assays discussed herein may, for instance, be used to detect such an antibody response. In some cases tolerance may be induced and an antibody response may not be seen and in particular subsequent challenge with the antigen may result in a lower response or no response in comparison to the response seen without the initial induction of tolerance.

The effects on other immune cells may be measured such as on macrophages and/or granulocytes. The skilled person will be aware of appropriate assays for assessing the immune response and immunogenicity.

In some embodiments of the invention a nucleic acid encoding an antigen, rather than the antigen itself, is targeted to the antigen presenting cell. In a particularly preferred embodiment the viral particles comprising targeting polypeptides may be employed in the invention. The ability of the viruses to selectively bind to antigen presenting cells may be measured. Non-antigen presenting cells may be employed as controls to assess the specificity of the viruses.

Standard techniques may be used to monitor the expression of the nucleic acids targeted to antigen presenting cells. For example, test experiments may be done using reported genes in place of an antigen encoding gene. Techniques such as RT-PCR, Northern blotting, Western Blotting and cell staining may be used to monitor expression in antigen presenting cells. Presentation of the antigen may be evaluated.

The above techniques may be used to identify effective targeting sequences. They may also be used to assess the efficacy of the invention in promoting an immune response against a chosen antigen. The techniques may also be used to assess the efficacy of the substances of the invention in the treatment or prevention of any of the conditions referred to herein. Typically suitable controls will be employed. In some cases standard test antigens and/or targeting polypeptides may be employed and compared to the substance under test.

#### *Loading antigen presenting cells*

The invention provides for the loading of antigen presenting cells. Thus the invention provides a method of loading an antigen presenting cell, comprising contacting an antigen presenting cell with a complex of the invention. The targeting polypeptide present in the complex directs the complex to the antigen presenting cells. The complex is taken up by the antigen presenting cell and the antigen is presented by the antigen presenting cell as discussed herein. The targeting of the complex to the antigen presenting cell is termed loading of the antigen presenting cell.

The invention also encompasses the loading of antigen presenting cells using a targeting polypeptide of the invention to deliver a nucleic acid encoding an antigen to antigen presenting cells. In particular viral particles comprising the targeting polypeptide may be employed. The invention therefore provides a method of loading an antigen presenting cell comprising using a targeting polypeptide of the invention to deliver a nucleic acid molecule encoding an antigen to the antigen presenting cell. The

invention provides for the infection of an antigen presenting cell with a virus of the invention.

The loading of antigen presenting cells or their precursors may be performed *in vitro*, *ex vivo* or *in vivo*. In the case of *in vitro* loading the antigen presenting cell may simply be contacted with a complex of the invention. The cell may be cultured in the presence of the complex under suitable conditions. The cell and complex may, for example, be contacted for between five minutes and ten days, preferably from an hour to five days, more preferably from five hours to two days and even more preferably from twelve hours to one day. *Ex vivo* loading may, for instance, be carried out in the same manner once the cells to be loaded have been obtained.

Loading of antigen presenting cells or their precursors may be performed *in vivo*. Thus the invention provides an *in vivo* method of loading antigen presenting cells comprising administering to a subject an effective amount of a complex of the invention. Administration may be via any of the routes discussed herein. Loading *in vivo* may also be achieved by administering a nucleic acid, vector, cell, virus, vaccine or pharmaceutical composition of the invention. The administration of such products results in complexes of the invention coming into contact with antigen presenting cells. Thus the nucleic, vector, virus or cells may lead to the production of a complex of the invention which then loads an antigen presentation cell present in the subject.

In one case antigen presenting cells recovered from a subject are loaded *in vitro* with a complex of the invention. The invention therefore provides an *ex vivo* method of loading antigen presenting cells. The loaded cells may then be returned to a subject and in particular to promote an immune response against the antigen that the cells have been loaded with.

The invention also provides antigen presenting cells which have been loaded using the complexes of the invention. Such antigen presenting cells will typically comprise the complex or breakdown products of the complex and in particular the antigen to be presented. The loaded cells may comprise epitopes of the antigen. The cell may also comprise the targeting polypeptide and/or breakdown products thereof. The invention provides such cells in an isolated form. The loaded cells may comprise a nucleic acid of the invention and in particular may comprise a viral vector of the invention. Preferably, the viral vector will be replication deficient.

Generally the antigen presenting cell of the invention carries peptides, and in particular an antigenic epitope, derived from the chosen antigen on its surface in

conjunction with an MHC class I or class II molecule and in particular in conjunction with an MHC II molecule. In one embodiment the antigen presenting cell has at least 100, preferably at least 200, for example at least or about 500 or 1000, class I and/or class II molecules on its surface loaded with the product and in particular class II molecules. In some cases, the cells may carry a label or be labelled, such as, for example, with thymidine or radioactive chromium. The invention also provides T cells that have been activated by a loaded antigen presentation cell of the invention.

In some cases antigen presenting cells may be recovered from a subject, loaded *in vitro* and then returned to the same subject. In other cases, it may be that cells are recovered from a subject, exposed to loaded antigen presenting cells of the invention *in vitro* and then returned to the subject.

In one case the invention may provide a composition comprising T cells, antigen presenting cells and a complex of the invention. The T cells or antigen presenting cells may be any of the cells mentioned above. In particular, the antigen presenting cells may be dendritic cells. The T cell:antigen presenting cell ratio may be typically from 500:1 to 1:500. Typically at least  $10^3$ , such as (e.g. at least or about)  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$  cells are present per millilitre of the composition. The composition typically also comprises a culture medium capable of supporting the T cells or antigen presenting cells, such as RPMI medium. The medium may also comprise cytokines, such as IL-2, IL-4, IL-7 or TNF- $\alpha$ . The T cells and antigen presenting cells may be from the same individual.

The cells employed in the invention, particular the antigen presenting cells and/or T cells, may be autologous cells, or cells which have been partially or fully matched with the subject for MHC class I HLA-A or HLA-B; and/or for MHC class II type. In a preferred case, the cells employed in the invention may be recovered from a subject and utilised *ex vivo* and subsequently returned to the same subject.

#### *Delivery of nucleic acid molecules encoding antigens to antigen presenting cells*

The targeting polypeptides of the invention may be used to deliver a nucleic acid molecule which encodes an antigen to an antigen presenting cell. The nucleic acid molecule gives rise to expression of the antigen in the antigen presenting cell and to the subsequent presentation of the antigen by the cell.

The nucleic acid molecule will preferably comprise a promoter which is operably linked to the sequences encoding the antigen and which is active in antigen presenting cells or which can be induced in antigen presenting cells.

5 In a particularly preferred embodiment the nucleic acid encoding the antigen may be delivered to the antigen presenting cell via a viral particle which comprises a targeting polypeptide of the invention. The targeting polypeptide will typically be provided wholly or partly on the surface of the virus in order for the polypeptide to be able to target the virus to an antigen presenting cell.

Any suitable virus may be used in such embodiments. The virus may, for  
10 example, be a retrovirus, a lentivirus, an adenovirus, an adeno-associated virus, a vaccinia virus or a herpes simplex virus. In a particularly preferred embodiment the virus may be a lentivirus. The lentivirus may be a modified HIV virus suitable for use in delivering genes. The lentivirus may be a SIV, FIV, or equine infectious anemia virus (EQIA) based vector. The virus may be a moloney murine leukaemia virus  
15 (MMLV).

The targeting polypeptide may comprise sequences from the virus. For example, the targeting polypeptide may comprise sequences from a viral surface protein. In particular, the targeting sequences may be at the N terminus of the targeting polypeptide and be fused or linked to surface protein sequences. The  
20 targeting polypeptide may also comprise a transmembrane domain so that it can be provided in the viral membrane.

In a particularly preferred embodiment the targeting polypeptide may include sequences from a surface protein of the virus. In particular the targeting polypeptide may comprise sequences from a surface protein that is involved in the normal binding  
25 of the virus to its target cell. The binding of the targeting polypeptide to the antigen presenting cell may lead to conformational changes allowing the viral surface protein sequences to bind to their target on the cell surface. This may lead to fusion of the viral and cell membranes allowing entry of the virus into the antigen presenting cell. Thus the targeting sequences and surface protein domains may show receptor  
30 cooperation to facilitate the entry of the virus specifically into the antigen presenting cells. The targeting polypeptide may include linker sequences which facilitate the cooperation and in particular proline rich sequences present in the viral surface protein may be employed. The proline rich linkers discussed in Martin *et al* (2003) Journal of

Virology 77(4): 2753-2756 and Valsesia-Wittmann *et al* (1997) EMBO Journal, 16(6): 1214-1223 may be employed.

The invention provides a virus comprising a targeting polypeptide as well as a cell infected by such a virus. The virus will typically also comprise a nucleic acid molecule encoding the chosen antigen. The virus may, for instance, be any of those mentioned herein. The nucleic acid molecule may also encode other sequences, for example the nucleic acid sequences may comprise sequences which express proteins which boost the immune response to the antigen. The nucleic acid may encode a cytokine, including any of those mentioned herein and in particular IL-1, IL2 and/or IL12. The nucleic acid may also encode a costimulatory molecule such as a surface polypeptide which enhances the immune response. The nucleic acid may encode, for example, CD80 and/or CD86.

The viruses of the invention are preferably replication deficient. In some cases the nucleic acid sequences encoding the targeting polypeptide will not be included in the viral vector. Thus the invention also provides helper cells which express the targeting polypeptide in such a way that the targeting polypeptide is incorporated into the viral particles. The invention also provides nucleic acid vectors that encode a targeting polypeptide of the invention which comprises viral sequences.

#### *Medicaments, methods and therapeutic use*

The complexes of the invention and various related aspects of the invention may be used in a method of therapy of the human or animal body. Thus the invention provides for the use of a targeting polypeptide, a complex, a nucleic acid, a vector, a cell, a virus, or an antigen presenting cell of the invention in a method of treatment of the human or animal body by therapy.

The invention provides for the use of a complex of the invention, a nucleic acid encoding a targeting polypeptide and antigen of a complex of the invention, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus of the invention or an antigen presenting cell of the invention in the manufacture of a medicament for use in immunisation.

In a preferred case the invention provides for the use of a complex comprising:

- (a) a targeting polypeptide comprising a staphylococcal superantigen-like

protein (SSL), a fragment thereof or a variant of either, where the SSL, fragment or variant has the ability to target the complex to an antigen presenting cell; and

(b) an antigen or a nucleic acid encoding an antigen,

5 in the manufacture of a medicament for use in immunisation.

In some instances the antigen comprises a polypeptide which is present in the complex as a fusion polypeptide with the targeting polypeptide. In others the antigen and targeting polypeptides are not part of the same polypeptide, but are covalently joined to each other or are joined through a linker. In a preferred case the antigen is a  
10 pathogenic antigen, an auto-antigen, an allergen and/or a cancer antigen. In another, the targeting polypeptide is present as a dimer.

Immunisation may result in promoting an immune response against the chosen antigen. Any of the effects resulting from targeting antigen presenting cells mentioned herein may be promoted or achieved. In particular the level of presentation  
15 of the chosen antigen will be increased. An increase in presentation via MHC I and/or MHC II molecules and in particular via MHC II molecules may be seen. In a preferred case the level of antigen presentation achieved may be such that when the same antigen is encountered again an increased immune response is seen in comparison to if the initial immunisation had not taken case. In particular a  
20 therapeutic and/or protective immune response may be raised. The invention may therefore ensure that a higher level of immune response is seen when the antigen is next encountered.

The invention may be used to enhance the level of antigen presentation or of any of the downstream effects thereof, such as any of those mentioned herein, in  
25 comparison to administration of an equivalent amount of antigen in the absence of a targeting polypeptide. The increase may be double, treble, or more fold, in some cases it may be at least ten-fold, preferably at least twenty-fold and even more preferably at least 100 fold, or 1000 fold or more. It may be that a protective, or therapeutic response, is seen whereas in the absence of the use of a targeting  
30 polypeptide it is not.

The invention also provides for an agent for immunising a subject, the agent comprising a complex of the invention, a nucleic acid encoding a targeting polypeptide and antigen of a complex of the invention, a vector comprising such a

nucleic acid, a cell comprising such a nucleic acid or vector, or an antigen presenting cell of the invention.

The various diseases and conditions to be prevented or treated may be any of those mentioned herein or associated with an antigen mentioned herein. In particular, the disease may be one associated with a pathogen, such as a bacterium, virus, bacterium, parasite, protozoan, fungus, and/or prion. In some cases the disease may be a cancer such as any of those mentioned herein.

In some instances the invention will be used to induce tolerance to a particular antigen and in particular to an allergen or an auto-antigen. In such cases typically the method will involve the delivery of the desired antigen to an antigen presenting cell in the absence of a stimulus which promotes antigen presenting cell maturation. In particular, the antigen may be delivered in the absence of an adjuvant such as aluminium hydroxide. The immunisation methods and vaccines of the invention may be used to induce tolerance to a selected antigen.

The substances of the invention may be administered to any suitable subject. The subject on which the method of the invention is performed is generally a vertebrate subject. By "vertebrate subject" is meant any member of the subphylum chordata, particularly mammals, including, without limitation, humans and other primates, as well as rodents, such as mice and rats. The subject may be a non-human animal. The non-human animal may be a domestic animal or an agriculturally important animal. The animal may be a domestic pet. The animal may be a monkey such as a non-human primate. The term subject does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. In one embodiment the subject is susceptible to or at risk from the relevant disease. For example, the subject may have been exposed, or will be in a region where there is a risk of exposure, to a particular antigen and in particular a pathogen.

The invention also covers the use of the complexes of the invention to promote antigen presentation of a chosen antigen. In some cases the invention may be used to bring about antigen presentation in an animal model, for example to study whether or not a particular immune response can be raised. The efficacy of the invention may be assessed in such non-human animal models. In some cases the invention may be used to help generate an immune response in a non-human animal in order to obtain antibodies against a chosen antigen that can then be recovered. Thus the invention may be used in a method of antibody production.

The invention may be used in combination with other means of, and substances for, immunisation. In some cases the complexes of the invention may be administered simultaneously, sequentially or separately with antigen which is not part of a complex of the invention. Thus complexes may be administered with the same antigen in a form not linked to a targeting polypeptide. The substances of the invention may be used in combination with existing vaccines for a particular antigen and may, for example, be simply mixed with such vaccines. Thus the invention may be used to increase the efficacy of existing vaccines including, for example, peptide, polypeptide, nucleic acid, viral and/or bacterial based antigens.

*Pharmaceutical compositions, vaccines and administration*

The invention additionally provides pharmaceutical compositions comprising a complex, nucleic acid, vector and/or cell of the invention and a pharmaceutically acceptable carrier or diluent. The present invention also provides a vaccine composition comprising a complex, nucleic acid, vector and/or cell of the invention. The vaccines and compositions may comprise any of the substances mentioned herein and in particular the complexes, nucleic acid molecules, vectors, viruses and cells of the invention. The invention provides a method of vaccination comprising administering to a subject an effective amount of a vaccine composition of the invention.

The various compositions, vaccines and other substances of the invention may be formulated using any suitable method. Formulation with standard pharmaceutically acceptable carriers and/or excipients may be carried out using routine methods in the pharmaceutical art. For example, an active substance may be dissolved in physiological saline or water for injections. The exact nature of a formulation will depend upon several factors including the particular substance to be administered and the desired route of administration. Suitable types of formulation are fully described in Remington's Pharmaceutical Sciences, 19<sup>th</sup> Edition, Mack Publishing Company, Eastern Pennsylvania, USA, the disclosure of which is included herein of its entirety by way of reference.

The substances may be administered by enteral or parenteral routes such as via oral, buccal, anal, pulmonary, intravenous, intra-arterial, intramuscular, intraperitoneal, topical or other appropriate administration routes. The substances may in some cases be administered to sites characterised by the presence of antigen

presenting cells. In cases where loaded antigen presenting cells are administered they may be administered, for example, to sites of antigen presentation such as secondary lymph nodes.

Vaccines may be prepared from one or more of the complexes, nucleic acids, vectors, and/ or cells of the invention together with a physiologically acceptable carrier or diluent. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the encapsulated in a liposome, particularly in the case of nucleic acids and vectors of the invention. The active ingredient may be mixed with an excipient which is pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, of the like and combinations thereof.

In addition, if desired, the vaccine and/or pharmaceutical compositions of the invention may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance effectiveness.

The complexes of the invention enhance the immunogenicity of a chosen antigen. They may therefore act as adjuvants for a chosen antigen and may be used as adjuvants. In some cases other adjuvants may be present in the various formulations of the invention or be administered simultaneously, separately or sequentially with them. Suitable adjuvants include, for example, any substance that enhances the immune response of the subject to the antigen (including when delivered by the polynucleotide of the invention). They may enhance the immune response by affecting any number of pathways, for example, by stabilizing the antigen/MHC complex, by causing more antigen/MHC complex to be present on the cell surface, by enhancing maturation of APCs, or by prolonging the life of APCs (e. g., inhibiting apoptosis).

Examples of adjuvants that may be employed include cytokines. Certain cytokines, for example TRANCE, flt-3L, and CD40L, enhance the immunostimulatory capacity of antigen presenting cells and may be employed. Non-limiting examples of cytokines which may be used alone or in combination include, interleukin-2 (IL-2), stem cell factor (SCF), interleukin 3 (IL-3), interleukin 6 (IL-6), interleukin 12 (IL-12), G-CSF, granulocyte macrophage-colony stimulating factor

(GM-CSF), interleukin-1 alpha (IL-1 a), interleukin-11 (IL-11), MIP-1a, leukemia inhibitory factor (LIF), c-kit ligand, thrombopoietin (TPO), CD40 ligand (CD40L), tumor necrosis factor-related activation-induced cytokine (TRANCE) and flt3 ligand (flt-3L). Further examples of adjuvants which may be effective include but are not limited to: aluminium hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion.

In cases where the invention is used to target a nucleic acid which encodes an antigen to an antigen presenting cell, the nucleic acid may also encode molecules capable of acting as an adjuvant. Thus the nucleic acid may lead to the production of any of the adjuvants mentioned herein and in particular a cytokine or costimulatory molecule. The cytokine may, for example be, IL-1, IL2, and/or IL-12 which will preferably be secreted from the antigen presenting cell. The costimulatory molecule may, for example, be CD80 or CD86 which will be preferably expressed on the cell surface of the antigen presenting cell.

The substances of the invention, and in particular, the vaccines, are typically administered parentally, by injection, for example, either subcutaneously or intramuscularly. Additional possible formulations include suppositories, oral formulations and formulations for transdermal administration. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1% to 2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10% to 95% of active ingredient, preferably 25% to 70%. Where the substance is lyophilised, the lyophilised material may be reconstituted prior to administration, e.g. a suspension. Reconstitution is preferably effected in buffer.

Capsules, tablets and pills for oral administration to a patient may be provided

with an enteric coating comprising, for example, Eudragit "S", Eudragit "L", cellulose acetate, cellulose acetate phthalate or hydroxypropylmethyl cellulose. Substances of the invention and in particular nucleic acids and vectors of the invention may be administered by needleless injection, for example, transdermally, may also be used.

5           The substances of the invention may be formulated as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salt (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric and maleic. Salts formed with the free carboxyl groups may also be  
10       derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine and procaine.

          The substances are administered in a manner compatible with the dosage formulation and in such amount will be prophylactically and/or therapeutically  
15       effective. The quantity to be administered depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgement of the practitioner and may be peculiar to each subject.

          A substance of the invention may be given in a single dose schedule, or  
20       preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of administration may be 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reinforce the immune response, for example at 1 to 4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be  
25       determined by the need of the individual and be dependent upon the judgement of the practitioner. Examples of dosages of complex will may be administered include from 5 µg to 100 mg, preferably from 50µg to 50 mg, more preferably from 250µg to 10 mg.

          In some cases the administered substances may comprise cells. The cells may,  
30       for example, be those comprising nucleic acids or vectors of the invention. In other cases the cells may be loaded antigen presenting cells or may be T cells that have had antigen presented to them by loaded antigen presenting cells of the invention. Any suitable number of cells may be administered to a subject. For example, at least, or about,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$  cells may be administered. As a guide the number of

cells of the invention to be administered may be from  $10^5$  to  $10^{13}$ , preferably from  $10^7$  to  $10^{11}$ . In such cases where cells are administered or present, culture medium may be present to facilitate the survival of the cells. In some cases the cells of the invention may be provided in frozen aliquots and substances such as DMSO may be present to  
5 facilitate survival during freezing. Such frozen cells will typically be thawed and then placed in a buffer or medium either for maintenance or for administration.

The nucleotide sequences of the invention and vectors can also be used administered as outlined above. Preferably, the nucleic acid, such as RNA or DNA, in particular DNA, is provided in the form of an expression vector, which may be  
10 expressed in the cells of the individual to be treated. The vaccines may comprise naked nucleotide sequences or be in combination with cationic lipids, polymers or targeting systems. The vaccines may be delivered by any available technique. For example, the nucleic acid may be introduced by needle injection, preferably intradermally, subcutaneously or intramuscularly. Alternatively, the nucleic acid may  
15 be delivered directly across the skin using a nucleic acid delivery device such as particle-mediated gene delivery. The nucleic acid may be administered topically to the skin, or to mucosal surfaces for example by intranasal, oral, intravaginal or intrarectal administration.

Uptake of nucleic acid constructs may be enhanced by several known  
20 transfection techniques, for example those including the use of transfection agents. Examples of these agents includes cationic agents, for example, calcium phosphate and DEAE-Dextran and lipofectants, for example, lipofectam and transfectam. The dosage of the nucleic acid to be administered can be altered. Typically the nucleic acid is administered in the range of 1pg to 1mg, preferably to 1pg to 10µg nucleic  
25 acid for particle mediated gene delivery and 10µg to 1mg for other routes.

The following Examples illustrate the invention.

## **Examples**

5

### **Example 1**

#### **Methods**

#### 10 *Recombinant protein expression and purification*

Recombinant N-terminal histidine tagged SSL7 and SSL9 proteins from *S. aureus* strain NCTC6571 were produced in *E. coli* using the expression vector pQE30 containing the genes NCTC6571*ssl7* and NCTC6571*ssl9* respectively as previously described (Williams,R.J. *et al.*, *Infect. Immun.* 68, 4407-4415 (2000)). Embp32 from  
15 *S. epidermidis* was also expressed as a recombinant N-terminal histidine tag fusion protein in *E. coli* and purified as previously described (Williams,R.J. *et al.*, *Infect. Immun.* 70, 6805-6810 (2002)).

#### *Crystallisation*

20 Crystals of SSL7 were obtained by the hanging drop vapour diffusion technique, at room temperature. Crystals were obtained in two different conditions. For the first condition (form I), the well buffer contained 25-30% (w/v) PEG-MME 2K, 0.2 M ammonium sulphate and 0.1 M MES, pH 6.5. Drops consisted of 1 µl recombinant SSL7 at 10 mg/ml and 1 µl well buffer. Crystals had a flat plate  
25 morphology with dimensions up to 0.3x0.2x0.01mm<sup>3</sup>. The second condition (form II) had the same protein concentration and drop size, but the well buffer in this case consisted of 28% (w/v) PEG 2K and 0.1 M Li<sub>2</sub>SO<sub>4</sub> buffered with 0.1M Tricine at pH 8.5. In this case the crystals were rod-shaped with approximate dimensions 0.3x0.05x0.05mm<sup>3</sup>.

30

#### *X-ray Data Collection*

Data were collected from cryocooled crystals following immersion in mother liquor containing 30% (v/v) glycerol (as described in Garmen,E. & Schneider,T. *Macromolecular Cryocrystallography. J. Appl. Cryst.* 30, 211-237 (1997)).

***Form I***

The data were collected to 2.75 Å resolution on line BM14 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France), on a Mar Research ccd  
5 detector. The data were indexed and integrated with Mosflm (Leslie, *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* 26, (1992)), and scaled and merged using Scala (Evans *et al.*, 97-103. 1997. CCLRC, Daresbury Laboratory. Ref Type: Conference Proceeding) from the CCP4 suite (*Acta Cryst.* **D50**, 760-763 (1994)). Subsequent analysis was carried out using programs from the CCP4 suite,  
10 unless otherwise stated. Data collection statistics are provided in Table 1.

**Table 1: Data Collection and Refinement**

Res. (Å)	N <sub>ref</sub>		R <sub>merge</sub> <sup>1</sup> (%)		I/σI		Completeness (%)		Redundancy (%)	
	Meas	Uni.	All	High <sup>2</sup>	All	High	All	High	All	High
Form I (P4 <sub>3</sub> 2 <sub>1</sub> 2)										
45.5- 2.75	59980	12607	9.4	46.0	14.6	3.1	98.7	99.2	4.5	4.3
Form II (P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> )										
30.0- 2.7	99169	11049	14.0	48.0	4.6	1.5	99.9	99.9	4.5	3.9

Protein		Water		R-factor <sup>3</sup> (%)		rmsd	
N	<sup>4</sup> B <sub>ave</sub> (Å <sup>2</sup> )	N	B <sub>ave</sub> (Å <sup>2</sup> )	Working <sup>5</sup>	Free <sup>6</sup>	Bonds(Å)	Angles (°)
Form I (P4 <sub>3</sub> 2 <sub>1</sub> 2)							
3127	43.6	18	30.9	23.5	27.5	0.008	1.35
Form II (P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> )							
	42.6	24	42.5	23.4	29.9	0.011	1.3

<sup>1</sup>  $R_{\text{merge}} = \sum |I_i - I_M| / \sum I_M$ , where  $I_i$  is the observed intensity of a reflection and  $I_M$  is the mean intensity of all related reflections.

<sup>2</sup> High: highest resolution shell. Form I: 2.9-2.75 Å; Form II: 2.85-2.70 Å.

<sup>3</sup>  $R\text{-factor} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ .

<sup>4</sup> Mean B-factor

<sup>5</sup> For the 95.1% of data included in the refinement.

<sup>6</sup> For the 4.9% of data randomly selected and excluded from refinement.

Autoindexing indicated that the crystals belonged to pointgroup P422, with cell dimensions  $a=b=81.66$  Å,  $c=148.04$  Å and a Matthews coefficient of 3.1 (corresponding to a solvent content of 60% v/v) for 2 molecules in the asymmetric unit. There was a significant peak in the native Patterson map at fractional coordinates: 0.000, 0.372, 0.500 and no peaks attributable to non-crystallographic symmetry in the self-rotation function, indicating the presence of two molecules related by a 2-fold rotation parallel to the a-axis in the asymmetric unit.

#### Form II

The data were collected to 2.7 Å resolution on line ID14-1 at the ESRF, on an ADSC Quantum 4R ccd detector. Data indexing, integration, scaling and analysis was carried out as for form I. Autoindexing and analysis of systematic absences indicated spacegroup P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, with  $a=51.65$  Å,  $b=71.59$  Å,  $c=103.47$  Å, and a Matthews' coefficient of 2.00, corresponding to a solvent content of 40%. There was a peak on the self-rotation function at  $\kappa=180^\circ$ , confirming the presence of two molecules in the asymmetric unit. Data collection statistics are provided in Table 1.

#### Molecular Replacement

The co-ordinates for one monomer of SSL9 (pdb-id: 1M4V; Arcus *et al.*, *J. Biol. Chem.* **277**, 32274-32281 (2002)) were used for molecular replacement, which was carried out with Molrep (Vagin & Teplyakov, *J. Appl. Cryst.* **30**, 1022-1025 (1997)).

#### Form I

There was a single clear peak in the rotation function ( $I/\sigma I=5.3$ , next peak  $I/\sigma I=3.58$ ). Since very few systematic absences were recorded during data collection, molecular replacement was carried out in all nine possible spacegroups to unambiguously identify screw axes. The best solution was for spacegroup P4<sub>3</sub>2<sub>1</sub>2, for which two molecules could be placed in the asymmetric unit, related by the appropriate translation, with a correlation coefficient of 32.6% (for comparison, the best solution in P4<sub>1</sub>2<sub>1</sub>2 had a correlation coefficient of 26.4%). Sidechains that differed between the SSL5 and SSL7 proteins were replaced by alanines in the

correctly positioned model, and rigid-body refinement was carried out using CNS version 1.1 (Brunger *et al.*, *Acta Cryst.* **D54**, 905-921 (1998)). At this point, the R-factor was 52.2% and the  $R_{\text{free}}$ , 54.6%. A round of simulated annealing reduced the R- and  $R_{\text{free}}$ -factors to 44.6 and 48.9% respectively.

5

### *Form II*

Molecular replacement was performed as described for form I. The best solution had a correlation coefficient of 37.4% for two molecules in the asymmetric unit. Following rigid-body and simulated annealing refinement in CNS, the R-factor and  $R_{\text{free}}$  were 34.16 and 42.06% respectively.

10

### *Density Modification*

For both crystal forms, cross-crystal averaging, non-crystallographic averaging and phase improvement were carried out using Dmmulti (Cowtan, *CCP4 Newsletter on Protein Crystallography* (1994)) in CCP4 (*Acta Cryst.* **D50**, 760-763 (1994)), prior to calculation of maps for manual rebuilding of the model.

15

### *Model Building and Refinement*

In both cases, manual rebuilding was performed using O (Jones & Kjølgaard, *Methods in Enzymology*. Charles W. Carter, J. & Sweet, R.M. (eds.), pp. 173-208 (Academic Press, 1997), and in later refinement rounds XtalView (McRae, *Practical Protein Crystallography*. Academic Press, San Diego, CA (1993)). Refinement was carried out with CNS. Alanine residues in the initial model were exchanged for the correct sidechains where positive Fourier difference density could be seen. At a number of positions the sequence alignment was incorrect and additional rebuilding of the chain was required. For the refinement, an overall anisotropic B-factor correction and bulk solvent scaling with  $k=0.36$ ,  $B=26.9 \text{ \AA}^2$ , for form I, and  $k=0.59$ ,  $B=101 \text{ \AA}^2$  for form II were applied. Noncrystallographic symmetry restraints were applied throughout refinement except in the later stages where there was clear evidence of a difference between the chains. After all protein residues had been included in refinement, a number of tightly bound waters were added, where there was a 3 rms peak in the difference Fourier, and a 1 rms peak in the  $2F_o - F_c$  map, and

20

25

30

appropriate protein-water hydrogen bonds. At the end of refinement, the R-factor was 23.5% and  $R_{\text{free}}$ , 27.5%, for form I and  $R=23.4\%$  and  $R_{\text{free}}=29.9\%$  for form II. A homology model of SSL9 was created using the program MODELLER (Lawkowski *et al.*, *J. Appl. Cryst* **26**, 283-291 (1993)) and both SSL7 and SSL5 as  
5 template structures.

Structural illustrations were drawn with Bobscript, ((Esnouf, *Acta Cryst.* **D55**, 938-940 (1999)) a modification of molscript (Kraulis, *J. Appl. Cryst.* **24**, 946-950 (1991) and rendered with Raster 3D (Merritt & Bacon, *Methods in Enzymology* **277**, 505-524 (1997) and Bacon & Anderson, *Journal of Molecular Graphics* **6**, 219-220  
10 (1988)).

#### *FITC labelling of SSLs*

SSL7 and SSL9 were dialysed against labelling buffer (0.2 M  $\text{NaHCO}_3$ , pH 9.0) overnight at room temperature (RT). 50  $\mu\text{l}$  of 1 mg/ml fluorescein  
15 isothiocyanate (FITC, Sigma) in dimethyl sulfoxide (DMSO) was added to 1 ml of a 2 mg/ml protein solution. After 4 hours incubation at room temperature in the dark, unbound FITC was removed by size exclusion chromatography using a PD-10 (Sephadex<sup>TM</sup>) column. The concentration of labelled protein, and the FITC:protein ratio were determined by spectrophotometry. All preparations gave FITC:protein  
20 ratios of between 1:1 and 2:1.

#### *Antibodies*

The following monoclonal antibodies (MAbs) were used: CD2 (mouse MAb MAS 593, IgG<sub>2b</sub>; Harlan), CD3 (supernatant mouse MAb UCHT1, IgG<sub>1</sub>; obtained  
25 from P. C. L. Beverley [Edward Jenner Institute for Vaccine Research, Compton, UK]), CD14 (supernatant mouse MAb HB246, IgG<sub>2b</sub>; gift from P. C. L. Beverley), and CD19 (supernatant mouse MAb BU12, IgG<sub>1</sub>; gift from D. Hardie [Birmingham University, Birmingham, UK]).

#### 30 *Cell culture*

Human PBMC-derived dendritic cells (DC) were generated from fresh whole blood samples obtained from healthy volunteers (Alderman *et al.*, *Cardiovasc. Res.*

55, 806-819 (2002) and Newton *et al.*, *Clin. Exp. Immunol.* 133, 50-58 (2003)). Mononuclear cells separated on Lymphoprep<sup>TM</sup> (Nycomed Pharma) by centrifugation at 400 g for 30 minutes were incubated in six-well tissue culture plates for 2 h at 37°C in 5% CO<sub>2</sub> in complete medium (CM)(RPMI 1640 medium (Gibco) supplemented with 10% fetal calf serum (FCS; PAA Laboratories), 100 U/ml penicillin, 100 µg/ml streptomycin, and 2 mM L-glutamine (Clare Hall Laboratories, Imperial Cancer Research Fund)). The adherent cells were cultured in fresh complete medium with 100 ng/ml human recombinant granulocyte-macrophage colony-stimulating factor (GM-CSF) and 50 ng/ml interleukin (IL)-4 (Schering-Plough Research Institute). On day four of incubation, loosely adherent cells were collected, and contaminating T and B lymphocytes were removed by incubation with CD3, CD2, and CD19 MAbs, followed by anti-mouse IgG-coated immunomagnetic Dynabeads<sup>TM</sup> (Dyna). The supernatant, containing highly purified DC was cultured for another three days in fresh complete medium with GM-CSF and IL-4. Human PBMC-derived macrophages were obtained using the same procedure for dendritic cell culture, except that 10% human serum was used and no cytokines were added (Swetman *et al.*, *Eur. J. Immunol.* 32, 2074-2083 (2002)).

#### *Binding and uptake of FITC labelled SSLs by human cells*

20 Binding assays were performed by incubating 10<sup>6</sup> cells/well in complete medium with various concentrations of SSL-FITC (0.05-1.25 µM) for 1 hour at 4°C or 37°C. In some experiments, 8 µM of unlabelled SSL was added to the cells together with the labelled protein. After incubation, cells were washed three times by centrifugation, and examined by flow cytometry. In some experiments, cells were additionally stained for various surface markers after SSL uptake. Cells were incubated with the relevant MAb for 30 min at 4°C, washed, and then incubated in 1: 25-diluted phycoerythrin-conjugated goat anti-mouse immunoglobulin (PE, Jackson ImmunoResearch) for 30 min at 4°C. Cells were washed, fixed in 2% formaldehyde and examined using a FACScan flow cytometer (Becton Dickinson).

30

### *Confocal microscopy*

10<sup>5</sup> cells were seeded on 32mm coverslips coated (for dendritic cells only) overnight at 4°C with 10µg/ml fibronectin (FN, Sigma) in HBSS (Gibco). After 2 hours at 37°C in complete medium, cells were incubated with SSL-FITC (1.25 µM) and/or Texas Red-dextran (1mg/ml, Molecular Probes) for 1 hour at 37°C in complete medium. The coverslips were then washed three times in cold HBSS and fixed in 2% paraformaldehyde. The slides were examined on a Bio-Rad Confocal Microscope. Images were acquired from 0.5-µm optical sections of individual cells.

## 10 **Results**

### *Structure determination of SSL*

Recombinant SSL7, consisting of residues 36-231 of the sequence with accession number AF094826 in GenBank, was crystallised in two different conditions each of which gave rise to a different crystal form.

#### *Form I*

The final, 2.75Å resolution, model built into the electron density map contained two SSL7 molecules, representing residues 18-213 of the recombinant SSL7. In addition, eighteen water molecules were included at stereochemically sensible locations. Though the His-Tag and N-terminal tail were disordered and omitted from the density, the majority of the residues had well-defined electron density. The final R-factor and R<sub>free</sub> were 23.5% and 27.5% respectively, and refinement statistics are given in Table 1. The final model had good stereochemistry, with 98.6% of residues in the most favoured and additionally allowed regions of the Ramachandran plot, and no residues in disallowed regions. The geometry was better than expected for the average 2.75 Å structure according to PROCHECK analyses (Lawkowski *et al.*, *J. Appl. Cryst* **26**, 283-291 (1993)).

#### 30 *Form II*

This model again contained two SSL7 molecules: in this case residues 21-213 or 23-213 of the construct in chains A and B respectively. Twenty-four water

molecules were added in this form. Once again, with the exception of the His-tag and N-terminal tail, the majority of the molecule had well-defined electron density. The final R- and R<sub>free</sub>-factors were 23.4 and 29.9% respectively. The final model had good stereochemistry with 96.5% of the residues in the two most favoured regions of the Ramachandran plot. The geometry was better than expected for the average 2.7 Å structure according to PROCHECK. Refinement statistics for both crystal forms are provided in Table 1.

### *The SSL7 structure*

The structure of the SSL7 monomer is shown in Figure 1(a). As predicted from sequence comparisons, the fold is similar to that of the bacterial superantigens, and consists of two domains. The N-terminal domain (residues 18-110) is an OB-fold, a variety of β-barrel associated with oligosaccharide and DNA binding, while the C-terminal domain (111-213) forms a β-grasp domain: a series of β-strands wrapped around a helix.

In total, the structure of four copies of the SSL7 monomer was obtained (two from each crystal form), and they are all very similar, as can be seen from the Cα-atom root mean square deviations (rmsds; Table 2).

**Table 2: Root Mean Square deviations (Å) between the different copies of the SSL7 monomer.**

		Form I		Form II	
		A	B	A	B
Form I	A	-	0.228	0.701	0.703
	B	-	-	0.669	0.642
Form II	A	-	-	-	0.450
	B	-	-	-	-

The only differences between the structures arise from differences in the conformations of flexible loops, and the linker between the N- and C-terminal

domain (residues 106-112). The relative orientation of the two domains to one another, and the orientations of the individual secondary structural elements remain unchanged in the different copies of the monomer. For this reason, unless explicitly stated, one example monomer, chain A from form I, will be used in the comparisons and discussions that follow.

### *SSL7 and other proteins*

When the SSL7 monomer is superposed on the structure of SSL7 (SET3) (the structure of SSL5 is provided by Arcus *et al.*, *J. Biol. Chem.* **277**, 32274-32281 (2002)), the other member of the family for which a three dimensional structure is available, the two are seen to share the same fold (Figure 1 (b)), as might be expected for proteins sharing 40% sequence identity. However, when optimally superposed, the rmsd, which is 1.33Å over 157 spatially equivalent C $\alpha$  atoms, is surprisingly high for two such highly related proteins (a value of about 1.1 Å over the whole structure would be anticipated from sequence identity (Chothia & Lesk, *EMBO J.* **5**, 823-6 (1986)).

The high rmsd can be largely accounted for by changes in two regions of the structure. Firstly, there is a change in the twist of the  $\beta$ -sheet in the C-terminal  $\beta$ -grasp domain (Figure 1, (b) right hand structure-indicated by arrow). The change in the twist results in shifts in individual residue positions as large as 6.65 Å (for the C $\alpha$  atom of G125 in the C-terminal domain). Secondly, there are changes to the conformations of the loops on the external face of the N-terminal OB-fold (Figure 1 (b), left hand structure- indicated by arrow), these loops are associated with a generic low affinity MHCII binding site in superantigens (Jardetzky *et al.*, *Nature* **368**, 711-8 (1994) and Kim *et al.*, *Science* **266**, 1870-1874 (1994)), and changes in them may indicate differences in function between the two proteins. These large movements account for the low contrast and large R-factor for the initial molecular replacement solution, prior to the simulated annealing which successfully realigned the sheet strands, and some of the loop residues.

The SSL7 structure was also compared to a homology model of SSL9, based on both SSL7 and SSL5 as template structures (not shown). The sequence of SSL9 has a much greater homology to SSL7 (sequence identity 49%) than SSL5 (sequence

identity 35%). This is reflected in the model of SSL9: when the structures are optimally superposed SSL9 has a C $\alpha$  atom rmsd of 0.6 Å over 188 spatially equivalent atoms from SSL7, and 1.4 Å over 177 spatially equivalent atoms from SSL5.

5           When SSL7, and the superantigen of known structure with which it shares the highest sequence identity (streptococcal pyrogenic exotoxin, SPEC; 29% - Roussel., *Nat. Struct. Biol.* 4, 635-43 (1997)) are optimally superposed (Figure 3), it is seen that once again the overall fold is conserved. SPEC has some extended loops, but the structures are otherwise very similar. Interestingly, despite the SPEC and SSL7  
10 sequences being far more divergent than SSL7 and SSL5, the two structures superpose nearly as well. The C $\alpha$  atom rmsd for optimally superposed SSL7 and SPEC is 1.48 Å over 134 structurally equivalent atoms. The difference between the structures on this occasion being both a slight change in the orientation of the  $\beta$ -grasp domain (Figure 1, (c) right hand structure, indicated by arrow), but also large  
15 differences in the orientations of the strands of the OB-fold (Figure 1, (c), left hand structure-indicated by arrow) the very large changes in conformation in this domain are not unexpected since the SSL proteins do not bind MHC II, and this is the region primarily involved in this interaction in the superantigens.

## 20    Dimerisation

In both crystal form I and II there are two molecules in the crystallographic asymmetric unit; these two molecules are related by a proper two-fold, resulting in the formation of an intimate dimer (Figure 1(d)). The dimer is virtually identical in both crystal forms, as can be seen from a comparison of the residues buried in the  
25 dimer interface (Figure 2), and the fact that the form I and form II dimers can be superposed with an all C $\alpha$  atom rmsd of 0.881 Å or 0.902 Å depending on orientation. The remainder of the crystal packing is entirely different for each of the forms, since the crystals grew in different conditions and at markedly different pH; it is unlikely that the dimers formed solely as a result of crystal packing forces.

30           As can be seen from Figure 1(d) the dimer interface is the result of the two  $\beta$ -grasp domains interacting to create an intermolecular  $\beta$ -sandwich. In the process, 1122 Å<sup>2</sup> of the monomer surface in form one and 1146 Å<sup>2</sup> of this surface in form II

are buried: this is in the range seen in biologically relevant dimers (Jones & Thornton, *Progress in Biophysics and Molecular Biology* 63, 31-59 (1995)). The dimer formation results in the burial of a number of hydrophobic residues (including F119, L128, I132) and a number of neutral polar residues contribute to a hydrogen bonding network between the two monomers, however, no charge-charge interactions are created. It has been shown that protein-protein interaction surfaces differ little from the 'normal' exterior surface of proteins, but that they tend to contain additional neutral polar residues and fewer charged ones (Lo Conte & Janin, *J. Mol. Biol.* 285, 2177-98 (1999)): the SSL7 dimer interface is entirely consistent with this. Figure 1 (d) also indicated with an arrow the loop with the largest difference between SSL7 and SSL5, the movement of which prevents steric clashes between the two SSL7 molecules in the dimer.

#### *Cellular tropism of SSL7 and SSL9*

Peripheral blood mononuclear cells (PBMC) were incubated with various concentrations of SSL7-FITC or SSL9-FITC for 1 hour, and cell-associated fluorescence measured by flow cytometry. Both SSL7 (Figure 3(a)) and SSL9 (not shown) stained a small proportion of PBMC at 37°C, but not at 4 °C. The level of fluorescence was dose-dependent up to a maximum at 1.25 µM protein. The mean percentage of cells stained with SSL7 (9.8±1.8, range 7.1-12.2, n=7) and SSL9 (10.9±1.1, range 9.4-12.6, n=5) was very similar. Mean fluorescence also increased with time between five minutes and a hundred and twenty minutes (not shown) suggesting progressive uptake of SSL protein by the cells.

In order to determine whether the interaction between SSL protein and PBMC was specific, competitive inhibition of SSL-FITC cell labelling by unlabelled SSL was investigated (Figure 3(b)). Excess unlabelled SSL7 was able to completely block uptake of SSL7-FITC. In contrast, neither SSL9, nor an unrelated bacterial protein also carrying a polyhistidine tag (Embp32) had any effect on the SSL7-FITC signal. Conversely, only unlabelled SSL9, but not SSL7 or Embp32, blocked uptake of SSL9-FITC. Interaction between SSL proteins and the PBMC therefore occurs via a saturable specific receptor, and is not mediated by the histidine tag on these

proteins. Furthermore, SSL7 and SSL9 use different receptors, or different sites within one receptor.

The PBMC sub-populations which are the targets for SSL7 and SSL9 were further characterised by immunophenotyping, using monoclonal antibodies to the major surface markers CD2, CD3, CD14, and CD19 (Figure 4). Both SSL7 and SSL9 were taken up by all CD14 positive cells, and by a population of CD2-low cells, a phenotype consistent with that of peripheral blood monocytes (Crawford, K. *et al.*, *J. Immunol.* **163**, 5920-5928 (1999)). Neither SSL7 nor SSL9 showed any interaction with CD3 positive T cells. Interestingly, SSL7-FITC but not SSL9-FITC stained a subpopulation of CD19 B cells, providing further evidence that the receptor for these two SSLs is distinct.

#### *Uptake of SSLs by dendritic cells*

Peripheral blood monocytes were cultured *in vitro* in the presence of GM-CSF and IL-4, in order to drive their differentiation into myeloid dendritic cells ((Sallusto & Lanzavecchia, *J. Exp. Med.* **179**, 1109-1118 (1994)). After depletion of residual lymphocytes, the population obtained after seven days culture consisted of >90% CD1a<sup>+</sup> HLA-DR high CD14 low dendritic cells (data not shown).

These cells were incubated for sixty minutes at 37°C with either SSL7-FITC or SSL9-FITC and examined by flow cytometry Figure 5 and confocal microscopy (data not shown). Dendritic cells stained uniformly strongly positive for both SSL7 and SSL9. Confocal microscopy confirmed that fluorescence was predominantly due to intracellular uptake of SSL, rather than surface staining. Both SSL7 and SSL9 were concentrated in small vesicular structures, localised particularly to the perinuclear region of the cell. In order to characterise the nature of these vesicles further, dendritic cells were cultured in the presence of SSL7 or SSL9-FITC and Texas Red dextran (data not shown), which is avidly taken up by dendritic cells via mannose receptors (Sallusto *et al.*, *J. Exp. Med.* **182**, 389-400 (1995)). Texas Red dextran strongly labelled a large number of intracellular vesicles throughout the dendritic cell cytoplasm. SSL distribution and dextran distribution partially overlapped, with some intracellular vesicles clearly containing both markers. However, SSL positive dextran negative vesicles were also observed. In a small

proportion of cells, vesicles containing SSL9 appeared to aggregate, to generate very large vesicles, which contained high concentrations of both SSL and dextran. The very large vesicles observed (which were never seen in the presence of dextran alone) presumably resulted from fusion of many SSL containing vesicles, and may  
5 have been driven by intramolecular interactions between SSL molecules (as observed during dimerisation in crystal structure). Similar vesicle distortion is observed in the presence of excess invariant chain, again driven by multiple interactions between invariant chain molecules (Romagnoli *et al.*, *J. Exp. Med.* 177, 583-596 (1993)).

In order to determine whether uptake of SSLs was a generalised feature of  
10 endocytic cells, peripheral blood monocytes were differentiated into macrophages, via culture in human serum, without added cytokines. Under these culture conditions, the cells develop a completely different phenotype (CD1a-, HLA-DR-, CD14 high) and morphology (lack of dendrite formation) (Swetman *et al.*, *Eur. J. Immunol.* 32, 2074-2083 (2002)). Macrophages, like dendritic cells efficiently endocytosed Texas  
15 Red dextran, but showed no uptake of either SSL7 or SSL9 (data not shown).

## Discussion

One of the most exciting results from the structural studies of SSL7 was the identification of an identical SSL7 homodimer in the asymmetric unit of two  
20 otherwise very different crystal forms, grown from very different solution conditions. The dimer has a number of characteristics seen in functionally relevant dimers, and which indicates that the dimer is not purely an artefact of crystallization. Interestingly, the structure of SSL5 did not reveal any such dimer formation (Arcus *et al.*, *J. Biol. Chem.* 277, 32274-32281 (2002)) and the residues making up the  
25 interface are not conserved across the different SSL proteins.

The change in the orientation of the  $\beta$ -grasp domain in SSL7 relative to SSL5 is necessary to allow the dimer to form: if a similar dimer is created from SSL5 monomers, clashes occur between residues 110-114 in one monomer and 197-200 in the other, and more seriously between residues 118-125 and 161-165. These clashes  
30 are alleviated by the change in the orientation of the  $\beta$ -grasp domain  $\beta$ -strands in SSL7, further suggesting that SSL7 does not form a dimer in the same way.

Since no crystal structure of SSL9 is yet available, a preliminary comparison with SSL7 was carried out using a homology model. As for SSL5, the residues involved in the dimer interface in SSL7 are not conserved between SSL7 and SSL9. However the sequence changes do not create steric or electrostatic clashes between the two monomers; rather they are such that hydrogen-bonding between the two molecules is maintained, and in some cases, new hydrogen-bonds are formed. SSL9 may therefore form a dimer in the same manner as SSL7.

SSL7 and SSL5 show differences in the region of the N-terminal domain that are implicated in a general low-affinity MHCII binding site in superantigens. The homology model reveals that the residues in these loops are in general highly conserved between the SSL7 and SSL9 sequences. However there are some important differences, including the change of P93 in SSL7 to threonine in SSL9. This proline is part of a well ordered  $\beta$ -turn in SSL7, while in SSL5 there is no ordered secondary structural element present here: in fact this in SSL5 loop is rather disordered. The pattern of sequence conservation in the N-terminal loops indicates that the structures of SSL7 and SSL9 are more related to one another than to SSL5 and this may also be reflected in the functional properties of the molecules. However, the small number non-conservative sequence changes between SSL7 and SSL9 are also entirely consistent with differences in their putative receptor binding, as discussed further below.

Studies of the superantigens have shown that their interactions with MHCII and T cell receptor molecules are diverse, encompassing a number of different interaction surfaces and stoichiometries. This includes the formation of functionally important superantigen dimers for some superantigens, for example the  $Zn^{2+}$ -dependent dimers formed by staphylococcal exotoxin D (Sundstrom *et al.*, *EMBO J.* **15**, 6832-40 (1996)), which form via the C-terminal  $\beta$ -grasp domain, in a manner reminiscent of the homodimers seen of SSL7. It also includes the formation of heterodimers using the same surface of the N-terminal OB-fold but different surfaces of MHC molecules (as in the complexes of HLA-DR1 with SEB and TSST-1 respectively - Jardetzky *et al.*, *Nature* **368**, 711-8 (1994) and Kim *et al.*, *Science* **266**, 1870-1874 (1994)).

The functional and structural studies described here show that SSL7 and SSL7, may be functionally active in different quaternary states.

The most significant differences between the structures of SSL7, SSL5 and other superantigens suggest that these molecules may interact with different binding  
5 partners, and this is supported by the studies of cellular tropism. The characteristic features of the interaction between SSL7 and SSL9, and PBMC are specificity, temperature dependence and cell selectivity. Specificity, indicative that the interaction is mediated by a cell surface receptor, is shown by the demonstration that unlabelled SSL blocks uptake of SSL-FITC. This competition is observed for both  
10 SSL7 and SSL9, ruling out the hypothesis that the results are due to significant differences in affinity of binding between the two.

The lack of reciprocal inhibition between SSL7 and SSL9 indicates that these two molecules have different binding partners on the cell surface, although the possibility that they bind to different sites on the same molecule cannot be ruled out.  
15 Although the binding sites of SSL9 and 5 on the cell surface are distinct, both are able to self-target to APCs. Since it was impossible to measure binding in the absence of uptake, true measurements of affinity could not be obtained. The concentrations required to obtain measurable uptake, however, were in the order of 0.1 micromolar, suggesting that the affinity of interaction with any putative receptor  
20 is relatively low. This is a characteristic of many classical superantigens (Labrecque *et al.*, *Semin. Immunol.* **5**, 23-32 (1993)).

The temperature and time dependence of SSL interaction are suggestive of receptor mediated uptake rather than simple binding to the cell surface, and this was confirmed by the confocal microscopy studies discussed further below. However, a  
25 small amount of surface binding can be detected at 37°C, but not 4°, using indirect labelling of intact cells with an antibody against the histidine tag (not shown). The interaction of SSL with the receptor, as well as its subsequent uptake, is therefore temperature-dependent.

The third characteristic of SSLs observed in these cellular studies with PBMC  
30 was the highly selective nature of the target population with which interaction could be detected. In *ex vivo* PBMC, the major target population is the monocyte, characterised by high expression of CD14. Essentially all monocytes were found to

interact with both SSL7 and SSL9. In contrast, neither SSL7 nor SSL9 interacted with T cells, identified by expression of CD3 and high levels of CD2. Interestingly, SSL7, but not SSL9 also bound to a proportion of B cells (in the order of 30% although this varied significantly between individuals), providing further evidence that the receptor for these two molecules is distinct. Since a very significant proportion of T cells, and all human B cells also express class II MHC (e.g. HLA-DR) this result rules out a direct binding of SSL7 or SSL9 to these molecules, thus clearly distinguishing them from classical superantigens.

Monocytes express both class I and class II MHC molecules, and can act as antigen presenting cells for the activation of CD4 or CD8 T cells. However, the prototype antigen presenting cell, and the only cell type which can activate naïve T cells, is the dendritic cell. It was therefore of interest that both SSL7 and SSL9 were taken up efficiently by monocyte-derived dendritic cell and hence both molecules can self target to this important class of antigen presenting cells. This cell type, which can be obtained by culture of PB monocytes in appropriate cytokines, provides a widely used model for myeloid dendritic cells. In contrast, neither SSL7 nor SSL9 showed any tropism for macrophages, a cell type also produced by *in vitro* culture of monocytes, but which has no antigen presenting capabilities.

Studies indicate that antigen presenting cell activity remains intact in the presence of SSLs. Conversely, self-targeting to antigen presenting cells results in enhancing the immunogenicity of these proteins. The uptake of SSL7 and SSL9 into an endosomal compartment which intersects with the dextran uptake pathway indicates that the SSL7 are successfully targeted to the antigen presentation pathway. This is because uptake via the mannose receptor efficiently targets antigens to the Class II MHC antigen processing pathway (Sallusto *et al.*, *J. Exp. Med.* **182**, 389-400 (1995)). Although enhancing immunogenicity would, at first sight, appear to be paradoxical, the generation of an antibody response to a secreted protein is unlikely to confer any advantage in bacterial clearance by the host. On the contrary, the interaction between secreted toxin and specific antibody in the microenvironment of the bacterium may activate complement and hence contribute to the breakdown of the physical barriers that restricts the invasiveness of these bacteria.

SSLs therefore appear to provide *S. aureus* with an alternative molecular strategy with which to distract the protective adaptive immune response of the host, and contribute to bacterial pathogenicity. Specifically, SSLs achieve this through their ability to target antigen presenting cells.

5

### **Example 2**

The work described in Example 1 shows SSLs (*staphylococcal* superantigen like proteins) interact selectively with antigen presenting cells, including dendritic  
10 cells. The functional consequences of this interaction are now examined further. We show that SSL uptake does not adversely effect any of the parameters of antigen presenting cell function examined using dendritic cells. SSL7 and 9 were found to have no effect on viability or morphology of dendritic cells. The proteins did not induce dendritic cell maturation, as measured by cell surface phenotype. Exposure to  
15 SSL did not alter the ability of dendritic cells to take up FITC-dextran. In addition, exposure to SSLs did not impair the ability of the dendritic cells to stimulate allogeneic or antigen specific T-cell responses.

The ability of antigen presenting cells to present SSLs was also examined. Dendritic cells loaded with SSL7 or 9 were able to stimulate a T-cell proliferative  
20 response in three out of eight healthy individuals tested. Sera from nine out of ten individuals tested contained antibodies against both SSL7 and SSL9, and the response to each SSL was specific and not cross-reactive.

The results obtained demonstrate that SSLs can be used to specifically target antigen presenting cells and gain access to the antigen presentation pathway of these  
25 cells. SSLs may therefore be utilised to specifically deliver chosen antigens to antigen presenting cells in order to elicit an immune response against the chosen antigen.

### **Methods**

30

### *Recombinant protein expression and purification*

Recombinant N-terminal histidine tagged SSL7 and SSL9 proteins from *S. aureus* strain NCTC6571 and Embp32 from *S. epidermidis* were produced as described above in Example 1. SSL proteins without histidine tag behaved  
5 identically to the tagged version in terms of cell binding and uptake (Al Shangiti AM *et al.*, Infect. Immun., 72:4261-70 (2004)).

### *Antibodies*

The following monoclonal antibodies (MAbs) were used: CD2 (mouse MAb  
10 MAS 593, IgG<sub>2b</sub>; Harlan), CD3 (supernatant mouse MAb UCHT1, IgG<sub>1</sub>; gift from P. C. L. Beverley [Edward Jenner Institute for Vaccine Research, Compton, UK]),  
CD14 (supernatant mouse MAb HB246, IgG<sub>2b</sub>; gift from P. C. L. Beverley), and  
CD19 (supernatant mouse MAb BU12, IgG<sub>1</sub>; gift from D. Hardie [Birmingham  
University, Birmingham, UK]), HLA-DR (supernatant mouse MAb L243, IgG<sub>2a</sub>; gift  
15 from P. C. L. Beverley), HLA-ABC (W6/32; Serotec), CD86 (supernatant mouse  
MAb BU63, IgG<sub>1</sub>; gift from D. Hardie), and CD54 (Mouse IgG MEM-111, gift from  
Prof. Horejsi, Academy of Science, Prague, Czech Republic). Fluorescein  
isothiocyanate (FITC) conjugated anti-mouse rabbit polyclonal antibody was  
purchased from Dako. PE-conjugated anti-CD1a (Monoclonal Mouse IgG1, clone  
20 BL6, Immunotech, Marseille, France). The rabbit polyclonal antibody to His-SSL7  
was produced under contract by Eurogentec Ltd (Southampton, UK) and validated  
by Western blot and ELISA.

### *Dendritic Cell Preparation*

25 Monocyte-derived human dendritic cells (MDDC) were generated from fresh  
whole blood samples obtained from healthy volunteers as described previously (Al  
Shangiti *et al.*, (2004), *supra*). Briefly, mononuclear cells separated on  
Lymphoprep<sup>TM</sup> (Nycomed Pharma) by centrifugation at 400g for 30 mins were  
incubated in 6-well tissue culture plates at 37°C in 5% CO<sub>2</sub> in complete medium  
30 (CM)(RPMI 1640 medium (Gibco) supplemented with 10% fetal calf serum (FCS;  
PAA Laboratories), 100 U/ml penicillin, 100 µg/ml streptomycin, 2 mM L-glutamine  
(Clare Hall Laboratories, Imperial Cancer Research Fund), 100 ng/ml human

recombinant granulocyte-macrophage colony-stimulating factor (GM-CSF) and 50 ng/ml interleukin (IL)-4 (Schering-Plough Research Institute). On day four of incubation, loosely adherent cells were collected, and contaminating T and B lymphocytes were removed by incubation with CD3, CD2, and CD19 MAbs, followed by anti-mouse IgG-coated immunomagnetic Dynabeads™ (Dyna). The non-adherent fraction, containing highly purified dendritic cells (less than 5% CD3, CD19 or CD14) was cultured for another three days in fresh culture medium with GM-CSF and IL-4.

#### 10 *Cell viability*

10<sup>5</sup> dendritic cells/group were cultured with SSL7 or SSL9 (4.16 µM) in a 96-well plate, and the plate was incubated at 37°C in 5% CO<sub>2</sub> overnight. Cell viability was assessed by trypan blue exclusion assay. 10 µl of cells suspension were diluted in an equal volume of trypan blue solution. 10 µl of this mix were loaded on a haemocytometer counting chamber placed under the microscope and white live cells (dead cells turn blue) were counted with in a 4 x 4 square grid.

#### *Cell surface phenotype expression*

Dendritic cell surface staining was performed by using a panel of monoclonal antibodies (MAbs) directed against surface antigens expressed by dendritic cells and the appropriate specific isotype controls. Briefly, 10<sup>5</sup> cells were pre-incubated for 24 hours in culture medium with 4.16 µM of SSL7 or SSL9 or with PG (5 µg/ml, from *S. aureus*, Sigma) or purified LPS (100 ng/ml; *Salmonella* Minnesota, Sigma) in 96 well U-bottomed plates at 37°C. Cells were resuspended in 100 µl of staining buffer (HBSS, 1% FBS, 0.1% sodium azide), and incubated first with the relevant MAbs for 30 mins at 4°C. Cells were washed, and secondary immunolabeling was performed using FITC-conjugated rabbit anti-mouse immunoglobulin (30 min, 4°C). Cells were washed three times and fixed in 3.8% paraformaldehyde and examined within 5 days on a FACScan flow cytometer (Becton Dickinson). Data were analysed using CellQuest software.

*Endocytosis assay*

Dendritic cells ( $10^5$ ) were incubated in culture medium with or with 4.16  $\mu$ M of SSL7 or SSL9 for various times (1 or 18 hours) in 96 well U-bottomed plates at 37°C. Different concentrations (1, 3, 10 and 30  $\mu$ g/ml) of FITC-dextran (40,000 MW) were incubated with the cells. After 1 hour of incubation at 37°C, cells were washed in ice cold HBSS containing 0.1% azide to stop further endocytosis, fixed with 3.7% formaldehyde, and analysed by flow cytometry. The uptake of dextran is expressed as mean fluorescent intensity. For each sample at least 5000 events gated on dendritic cells were analysed.

*T-cell Proliferation Assays*

Autologous T-cells were obtained from non-adherent population of peripheral blood mononuclear cell fraction from eight healthy volunteers (age range 20-50, median approximately 30) and Cryopreserved in FCS containing 10% DMSO (Sigma Aldrich) at -70°C. Cells were thawed rapidly (37°C), and B cells, monocytes, and macrophages were depleted by incubation with CD19, HLA-DR and CD14 MoAb for 45 minutes on ice. Cells were washed and then mixed with magnetic microbeads and separated on magnetic columns. T-cells (greater than 90% purity) were used immediately after purification. Allogeneic T-cells used in mixed leucocyte reactions (MLR) were prepared from HLA-mismatched donors in same procedure. Purified dendritic cells ( $10^4$ ), either untreated or treated for 18 hours to different concentration of SSL proteins (4.16, 1.25 and 0.42  $\mu$ M), were incubated at 37°C/5% CO<sub>2</sub> with autologous T-cells ( $2 \times 10^5$  cells/well) in the presence of purified protein derivative (PPD) or with allogeneic T-cells in flat-bottomed 96-well microtiter plates. The dendritic cell autologous and allogeneic T-cell cocultures were incubated 6 days. Both assays were then pulsed with 1  $\mu$ Ci of [<sup>3</sup>H]thymidine (ICN Biomedical, High Wycombe, United Kingdom) for the final 18 hours of culture. Cells were harvested, and T-cell proliferation was measured by liquid scintillation counting (Microbeta Systems). All assays were performed in triplicate. Results were express as cpm. Error bars represent the standard deviation (SD).

*Cytokines assays*

Autologous T and dendritic cells were incubated with different concentrations of SSL proteins (4.16, 1.25 and 0.42  $\mu$ M) at 37°C/5% CO<sub>2</sub> in 24-well plates. After 4 days, cell culture supernatants were centrifuged for the removal of cells and stored at  
5 -70°C. Cytokine detection was done by enzyme-linked immunosorbent assay (ELISA) for interleukin-10 (IL-10, Pharmingen, UK), gamma interferon (IFN- $\gamma$ , Pharmingen, UK) and IL-13 (ImmunoTools, Germany). Purified protein derivative (PPD; 500 U/ml) was used as positive control.

10 *Antibody detection*

Human serum was collected from heparinised blood of ten normal individuals (three females, seven males, age range 20-50, median 30). ELISA microassay plates were coated for 24 hours with 0.04  $\mu$ M of SSL proteins dissolved in sodium carbonate buffer 0.1 M, pH = 9.5 at 4°C (100  $\mu$ l/well). After three successive washes  
15 with HBSS containing 0.1% Tween 20<sup>TM</sup>, blocking was performed with 1% skimmed milk in HBSS for 1 hour at 37°C. The plates were washed again three times, as before, and the tested sera diluted at 1:2000 in HBSS with 0.1% Tween<sup>TM</sup> were added to the SSL coated plates for 1 hour at 37°C (preliminary studies showed that this dilution of antisera gave no background staining for any serum tested in  
20 control wells). After three additional washes, the remaining bound antibodies were incubated for 1 hour at 37°C with alkaline phosphates – conjugated human antibodies diluted at 1:1000 in HBSS with 0.1% Tween 20<sup>TM</sup>. Excess conjugate was removed by washing as above and a colorimetric reaction was carried out by addition of the chromogen OPD (o-Phenylenediamine dihydrochloride) for 15-20 minutes. The  
25 plates were read (405 nm) to detect the optical density (OD) readings. A control well containing no serum was used to detect the background count. A polyclonal rabbit antibody raised against purified His-tagged SSL was included as a positive control (anti-SSL7).

Competitive ELISA was performed by mixing the sera with differing  
30 concentrations of SSL or a control bacterial protein Embp32 (0.08, 0.17, 0.33 and 0.42  $\mu$ M), and then testing binding on SSL-coated plates as above.

*Statistical analysis*

The means of paired groups were analysed by a 2-tailed Student's *t* test. The level of significance was <0.05.

## 5 Results

*The effects of SSL protein on DC viability, morphology and surface phenotype*

Monocyte-derived dendritic cells (MDDC) were derived from peripheral blood monocytes. They showed the characteristic phenotype with low CD14, and high HLA-DR and high CD1a. As shown in Figure 1, fluoresceinated SSL7 and SSL9 labelled CD1a expressing cells dendritic cells in unpurified dendritic cell cultures, while residual contaminating cells (predominantly lymphocytes) did not show any interaction with either protein. Nevertheless, in order to exclude the possibility of indirect effects mediated on dendritic cells via some other cell type all further experiments were performed on purified dendritic cell cultures, containing less than 5% non-dendritic cells.

Dendritic cells were incubated with SSL7 or SSL9 (4.16  $\mu$ M) for either 1 or 18 hours at 37°C and cell viability was assessed by trypan blue exclusion. Microscopic analysis revealed that SSL proteins were not cytotoxic as more than 95% of cells appeared viable. Untreated cells were predominantly non-adherent with few dendritic cell processes, typical of immature dendritic cells (Figure 7A, top panel). Neither SSL7 or SSL9 induced any noticeable morphological changes over the time period tested (Figure 7A, middle panels). In contrast dendritic cells treated with the TLR4 bacterial ligand LPS (100 ng/ml) or peptidoglycan (PG) 5  $\mu$ g/ml became adherent and extended multiple, long dendritic processes (Figure 7A, bottom panels). The cell-surface expression of a panel of characteristic dendritic cell surface markers was analyzed by flow cytometry. Immature dendritic cells were incubated with 4  $\mu$ M SSL7 or SSL9, or LPS and PG and the surface phenotype of these dendritic cells was analysed after 18 hours of culture. Neither SSL7 or SSL9 (Figure 7B) induced significant changes in any of the surface molecules measured. In contrast, dendritic cells incubated with either LPS or PG up-regulated surface expression of HLA-DR, HLA-ABC, CD86 and CD54. Thus, in summary, exposure

of dendritic cells to the SSLs protein did not induce dendritic cell maturation, nor indeed any obvious changes in dendritic cell surface phenotype, viability or morphology.

5     *The influence of SSLs on endocytosis*

Fluorescein isothiocyanate-labelled dextran (FITC-Dx) is rapidly taken up by dendritic cells via the mannose receptor (Sallusto F., *et al.*, J. Exp. Med., 182:389-400 (1995)). To determine whether SSL protein altered antigen uptake function, dendritic cells were treated with 4.16  $\mu$ M of SSL7 or SSL9 and incubated for 1 or 18  
10     hours at 37°C. Different concentrations of FITC-Dx (1, 3, 10 and 30  $\mu$ g/ml) were added to the cell and incubated for a further hour at 37°C, and FITC-Dx uptake by the dendritic cell was then measured by flow cytometry. Figure 8 shows the results obtained with the total cell associated dextran being measured by flow cytometry and expressed as mean fluorescent intensity for a minimum of 5000 dendritic cells. As  
15     shown in Figure 8, SSL treated dendritic cell showed rapid uptake of FITC-Dx, and neither protein had any effect on endocytic activity.

*The influence of SSLs on the T-cell stimulatory capacity of dendritic cells*

To evaluate the effect of SSL proteins on the stimulatory capacity of dendritic  
20     cells in T-cell proliferation, day 6 dendritic cells ( $10^4$ ) were incubated for 18 hours with different concentrations of SSL proteins (0.42, 1.25, and 4.16  $\mu$ M). Residual T-cells were depleted and the functional assays performed using fresh viable purified autologous or allogeneic T-cells ( $2 \times 10^5$ ). The ability to induce secondary immune responses was unchanged. Figure 9A shows representative experiments eliciting  
25     recall responses to tuberculin (PPD – 500 U/ml)). There was no statistical difference in the proliferative responses observed between any of the pre-incubated dendritic cell groups with SSL 7 or SSL9 ( $37182 \pm 2036$  cpm and  $36458 \pm 3000$  cpm, respectively) and the control ( $36458 \pm 6151$  cpm) after the 6 days co-culture period ( $P > 0.05$ ).

30     The capacity of SSL-treated dendritic cells ( $10^4$ ) to elicit primary T-cell proliferation also was tested in an allogeneic mixed lymphocyte reaction (MLR). The same concentrations of SSL 7 and 9 and number of T cells used in the autologous

assay were employed with the allogeneic T cells. A similar result was observed (Figure 9B), as proliferation response against allogeneic T-cells ( $25710 \pm 1140$  cpm) was unaffected by the treatment of SSL7 or SSL9 ( $26149 \pm 3674$  and  $25816 \pm 3159$  cpm, respectively ( $P > 0.05$ )). Therefore, SSL proteins have no effect on the ability to induce proliferation of allogeneic T-cells. In general, the results of these experiments demonstrate that the antigen presentation capacity of dendritic cells remains intact in the presence of these secreted proteins.

#### *T-cell responses to SSLs in the normal human population*

To investigate the ability of SSLs to stimulate a recall T-cell response in healthy volunteers, dendritic cells ( $10^4$ ) were incubated with autologous T-cells ( $2 \times 10^5$ ) from normal donors in the presence of different concentrations of SSL proteins (0.42, 1.25, and 4.16  $\mu$ M) for 6 days (Figure 10). Purified protein derivative (PPD) was used as a positive control. A recall response against SSL7 was documented in 2/8 individuals and a response to SSL9 in 3/8 individuals. All volunteers showed a good recall response against PPD ( $70889 \pm 3146$  rpm).

The supernatants of three dendritic cell/T-cell/SSL co-cultures (individuals 1, 2 and 8 from Figure 10) were tested for IFN- $\gamma$  (TH1), IL-13 (TH2) and IL-10 (Treg) after 4 days. All cytokine levels were low (IFN $\gamma$  < 700pg/ml n=3); IL-13 < 50 pg/ml, n=3) or undetectable (IL-10). The results for the individual with maximum response (individual 1 in Figure 10) at different SSL concentrations are shown in Figure 11.

#### *Antibodies responses to SSLs in the normal human population*

In order to see if the presence of a T-cell response correlated with antibody production, sera from ten individuals (including those tested for T-cell responses as shown above) were tested by ELISA against immobilized SSL7 and SSL9 (Figure 12A). Sera was diluted 1:2000 and tested for binding to SSL7 or SSL9 by ELISA as described in the materials and methods section. Nine out of ten individuals tested showed antibody responses to both SSL7 and SSL9 at this dilution. Interestingly, competitive ELISA (Figure 12B and C) showed that the antibody response was highly specific for individual SSL isotypes. Increasing concentrations of SSL7

(Figure 12B) were able to completely block the interaction between SSL7 protein and SSL7 sera. In contrast, neither SSL9, nor an unrelated bacterial protein (Embp32) had any effect on the SSL7 antibody binding. Conversely, only SSL9, but not SSL7 or Embp32, were able to block SSL9-antibody (Figure 12C). Therefore,  
5 interaction between SSL proteins and SSL-antibodies were specific and do not cross-react.

## Discussion

10

Dendritic cells as professional antigen presenting cells have a key role in the initiation of the immune response against microbial infections; therefore, many microbial strategies have been described which interfere with dendritic cell function (Moll H., Cell Microbiol., 5:493-500 (2003)). One possibility was that SSLs might  
15 interfere with normal function of dendritic cells and therefore impair the protective immune response to *S. aureus*. Such functions have recently been proposed for the anthrax lethal toxin (Agrawal A, *et al.*, Nature, 424:329-34 (2003)) and *E. coli* heat labile toxin (Petrovska L, *et al.*, Vaccine, 21:1445-54 (2003)). The possibility of SSLs inhibiting dendritic cell function was therefore ruled out in the present study.

20 SSL7 or SSL9 were shown to be non-toxic to antigen presenting cells and did not alter the characteristic morphology of these cells (cf the effect of *Clostridium difficile* toxin B, (Swetman CA *et al.*, Eur. J. Immunol., 32:2074-83 (2002)). Conversely, SSL7 and SSL9 did not induce process extension, or up-regulation of cell surface co-stimulatory and HLA molecules on the dendritic cell, two  
25 characteristic signs of activation/maturation responses induced by whole *S. aureus* (Tourkova IL, *et al.*, Immunol. Lett., 78:75-82 (2001)) or bacterial surface components such as peptidoglycan (PG) (Michelsen KS, *et al.*, J. Biol. Chem., 276:25680-6 (2001)). Thus, although SSLs bind to and are taken up by dendritic cells (Figure 6) this interaction does not appear to engage activating receptors on the  
30 dendritic cell surface. Previous studies (Williams RJ, *et al.*, Infect. Immun., 68:4407-15 (2000)) indicating that SSLs could activate high levels of inflammatory cytokine release from peripheral blood cells could not be repeated using the highly

purified protein preparations used for this study (data not shown) and may have resulted from trace amounts of contaminating LPS.

In addition to their specialized dendritic morphology and cell surface phenotype, dendritic cells are characterized by extremely rapid endocytosis by both fluid phase and receptor mediated uptake (Swanson JA, *et al.*, Trends Cell Biol., 5:424-8 (1995) and Levine TP, *et al.*, Adv. Exp. Med. Biol. 329:11-5 (1993)). The uptake of FITC-Dextran, which is believed to be mediated via mannose receptors on the cell surface (Sallusto F, *et al.*, J. Exp. Med., 182:389-400 (1995)) is frequently used to measure the latter. Dendritic cells did indeed show efficient internalization of FITC-Dextran (albeit slightly less well after overnight culture) and this uptake was not altered by exposure to SSL7 or SSL9. Finally, since dendritic cells are distinguished by being the most potent stimulators of both primary and secondary T-cell responses, we tested the effects of SSL7 and SSL9 exposure on dendritic cell function directly. Although dendritic cells stimulated powerful proliferative responses to both PPD (a classical recall secondary response to BCG vaccination) and allogeneic purified T-cells (predominantly a primary response) neither SSL7 nor SSL9 altered the antigen presentation activity of dendritic cells. Taken together, therefore, these data do not provide any evidence that SSL proteins inhibit or modify dendritic cell function.

Further experiments demonstrated that SSL targeted to antigen presenting cells, in particular dendritic cells, actually are delivered to the antigen presentation pathway. This can hence enhance an immune response to these proteins and can also be used to deliver chosen antigens to the same pathway. The immune response to SSLs was analysed in a small panel of healthy human volunteers. Although none of the individuals tested had any known history of clinical *S. aureus* infection, the organism is extremely prevalent in the environment and approximately 30-40% individuals are persistently colonized by *S. aureus*, usually in the nasal mucosa (Nair SP, Williams RJ, Henderson B, Advances in our understanding of the bone and joint pathology caused by *Staphylococcus aureus* infection, Rheumatology, (Oxford), 2000; 39:821-34). Indeed, in eight volunteers tested, three (37%) showed a dose dependent T-cell response to dendritic cells loaded with SSL9 and two to SSL7. The response was detectable with relatively large numbers of T-cells/well ( $2 \times 10^5$ ) and

induced the release of a very low level of either TH1 (IFN- $\gamma$ ) or TH2 (IL-13) cytokines, suggesting precursor frequency of T-cell specific for SSL was likely to be low.

Despite a low T-cell precursor frequency, the humoral response to SSL7 and  
5 SSL9 was robust. Using solid phase ELISA nine out of ten sera tested showed specific antibody binding to both SSLs. The response measured was IgG (using an anti-IgG detection antibody) suggesting that class switching had occurred and further implicating the activity of SSL-specific T-cells. Interestingly, some individuals (e.g. individuals 2 and 7 in Figures 10 and 11) show antibody responses, but no detectable  
10 T-cells responses, perhaps because precursor T-cell frequency has fall below detectable levels in these individuals. One individual showed a T-cell response to SSL9, but no antibody response to either SSL tested, although we cannot rule out that some antibody might be detectable at a lower dilutions. Interestingly, the antibody response to each SSL was highly specific with minimal evidence of cross-  
15 SSL reactivity. This data is consistent with the sequence diversity between SSL paralogs, despite a highly conserved three dimension structure (Al Shangiti *et al* (2004) *Supra* and Arcus VL, *et al.*, J. Biol. Chem., 277:32274-81 (2002)). The presence of SSL specific immunity in so many individuals, and the existence of so many SSL paralogs in the *S. aureus* genome, is suggestive of a strong evolutionary  
20 interaction between host immunity and this bacterial family of proteins.

In conclusion, this study describes a number of functional consequences of the interaction between SSL and dendritic cells. In contrast to some other bacterial exotoxins (Agrawal *et al*, (2003), *Supra* and Petrovska *et al.*, (2003) *Supra*) SSLs do not appear to damage dendritic cells, but rather can be taken up by them, and thus  
25 stimulate T-cell response in healthy individuals. This means that SSLs may be employed to selectively deliver chosen antigens to antigen presenting cells, in particular to dendritic cells. SSLs may therefore be used to help induce an immune response or tolerance to a selected antigen.

30

CLAIMS

1. Use of a complex comprising:
  - 5 (a) a targeting polypeptide comprising a staphylococcal superantigen-like protein (SSL), a fragment thereof or a variant of either, where the SSL, fragment or variant has the ability to target the complex to an antigen presenting cell; and
  - (b) an antigen and/or a nucleic acid molecule encoding an antigen,
- 10 in the manufacture of a medicament for use in immunization or the induction of tolerance.
2. Use according to claim 1, wherein the antigen comprises a polypeptide which is present in the complex as a fusion polypeptide with the targeting polypeptide.
- 15 3. Use according to Claim 1 or 2, wherein the antigen and targeting polypeptides are not part of the same polypeptide, but are covalently joined to each other or are joined through a linker.
- 20 4. Use according to any one of the preceding claims, wherein the antigen is a pathogenic antigen, an auto-antigen, an allergen and/or a cancer antigen.
5. Use according to any one of the preceding claims, wherein the targeting polypeptide is present as a dimer.
- 25 6. Use according to any one of the preceding claims, wherein the targeting polypeptide comprises:
  - (a) a polypeptide having the amino acid sequence of any of SEQ ID Nos 6, 7, 9, 20, 21, 23, 30, 32, 34, 40, 44, 42, 58, 59, 60 72, 74, and/or 76;
  - 30 (b) a fragment of any of the sequences of (a), the fragment having the ability to target the complex to an antigen presenting cell; and/or
  - (c) a variant polypeptide having at least 30% amino acid sequence identity

to any of the polypeptides of (a) or (b) and the ability to target the complex to an antigen presenting cell.

7. Use according to claim 6, wherein the targeting polypeptide comprises:
- 5 (a) the sequence of SEQ ID No: 7, 9, 21, 23, 32, 34, 40, 42, 59, 60, 74, 76 and/or 92;
- (b) a fragment of any of the sequences of (a), the fragment having the ability to target the complex to an antigen presenting cell; and/or
- (c) a variant polypeptide having at least 70 % amino acid sequence identity  
10 to any of the polypeptides of (a) or (b) and the ability to target the complex to an antigen presenting cell.

8. Use according to any one of the preceding claims wherein the medicament is for the induction of tolerance and is to be administered without an adjuvant.

15

- 9 A complex comprising:
- (i) a targeting polypeptide as defined in any one of the preceding claims; and
- (ii) an antigen or a nucleic acid encoding an antigen, wherein the antigen or encoded antigen is selected from a pathogenic antigen, auto-antigen, an allergen and  
20 a cancer antigen.

10. A complex according to claim 9, wherein the targeting polypeptide is present as a dimer.

- 25 11 A virus comprising a targeting polypeptide as defined in any one of claims 1 to 8.

12. A nucleic acid molecule comprising a polynucleotide sequence encoding a targeting polypeptide and antigen, wherein the targeting polypeptide and antigen are  
30 as defined in claim 9.

13. A nucleic acid according to claim 12, wherein the nucleotide sequence encoding the targeting polypeptide and the antigen are present in a single open reading frame.
- 5 14. A vector comprising a nucleic acid according to claim 12 or 13.
15. A vector according to claim 14, comprising a promoter capable of giving rise to expression of both the targeting polypeptide and the antigen in an antigen presenting cell.
- 10 16. A cell comprising a nucleic acid according to claim 12 or 13 or a vector according to claim 14 or 15 or infected with a virus according to claim 11.
- 15 17. A method of loading antigen presenting cells comprising contacting an antigen presenting cell or a precursor thereof with a complex as defined in any one of claims 1 to 10 or a virus according to claim 11
18. A method according to claim 17, which is an *in vitro* method.
- 20 19. An antigen presenting cell which has been loaded with a complex as defined in any one of claims 1 to 10 or a virus according to claim 11.
- 25 20. A pharmaceutical composition comprising a complex as defined in any one of claims 1 to 10, a nucleic acid encoding the targeting polypeptide and antigen of a complex as defined in any one of claims 1 to 10, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus according to claim 11 or an antigen presenting cell according to claim 19 and a pharmaceutically acceptable carrier or diluent.
- 30 21. A vaccine comprising a complex as a complex as defined in any one of claims 1 to 10, a nucleic acid encoding the targeting polypeptide and antigen of a complex as defined in any one of claims 1 to 10, a vector comprising such a nucleic

acid, a cell comprising such a nucleic acid or vector, a virus according to claim 11 or an antigen presenting cell according to claim 19.

22. A complex as defined in any one of claims 1 to 10, a nucleic acid encoding the targeting polypeptide and antigen of a complex as defined in any one of claims 1 to 10, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus according to claim 11, or an antigen presenting cell according to claim 19 for use in a method of treatment of the human or animal body by therapy.

23. Use of a nucleic acid encoding the targeting polypeptide and antigen of a complex as defined in any one of claims 1 to 10, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus according to claim 11 or an antigen presenting cell according to claim 19 in the manufacture of a medicament for use in immunisation.

24. A method of immunising a subject, the method comprising administering an effective amount of a complex as defined in any one of claims 1 to 10, a nucleic acid encoding the targeting polypeptide and antigen of a complex as defined in any one of claims 1 to 10, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, a virus according to claim 11 or an antigen presenting cell according to claim 18 to a subject.

25. An agent for immunising a subject, the agent comprising a complex as defined in any one of claims 1 to 9, a nucleic acid encoding the targeting polypeptide and antigen of a complex as defined in any one of claims 1 to 9, a vector comprising such a nucleic acid, a cell comprising such a nucleic acid or vector, or an antigen presenting cell according to claim 17.

30

1/14

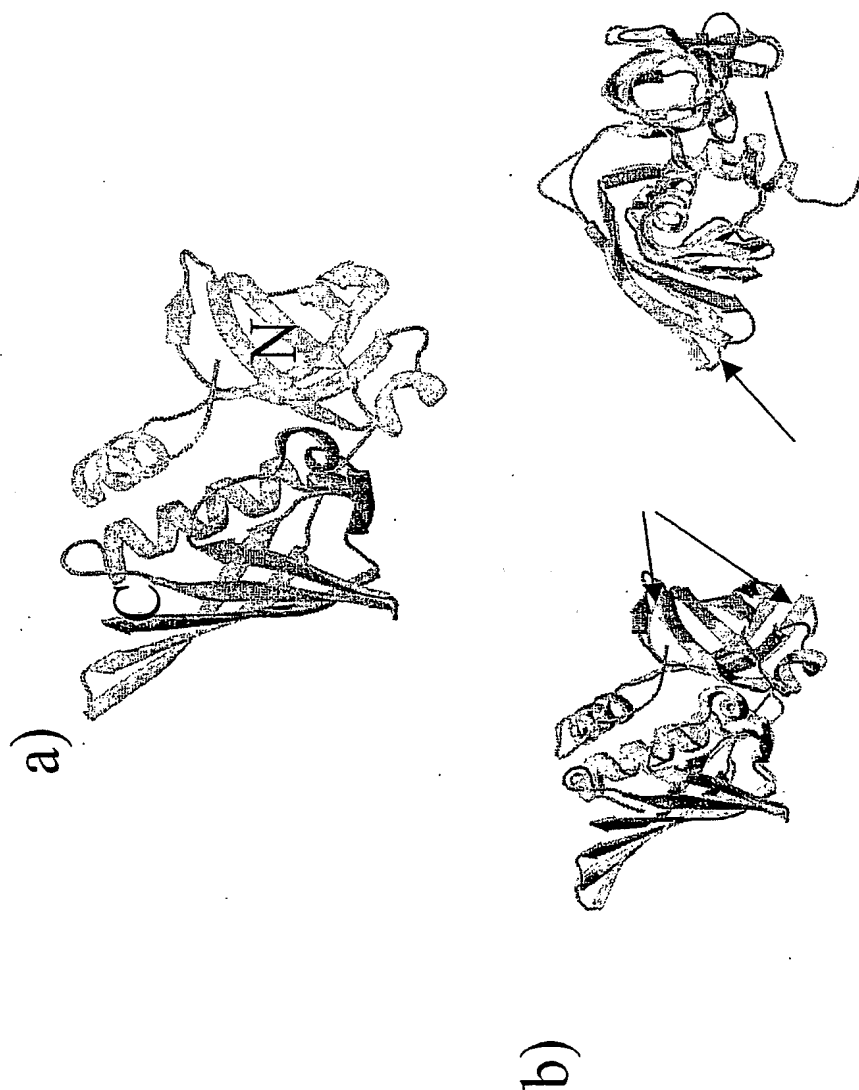


Figure 1

2/14



Figure 1 (continued)

3/14

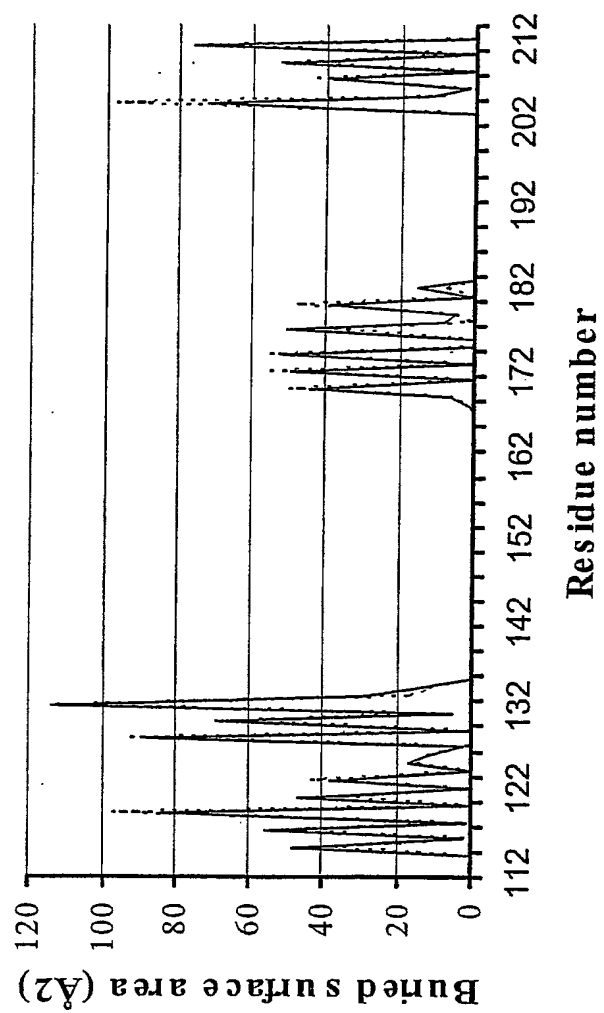


Figure 2

4/14

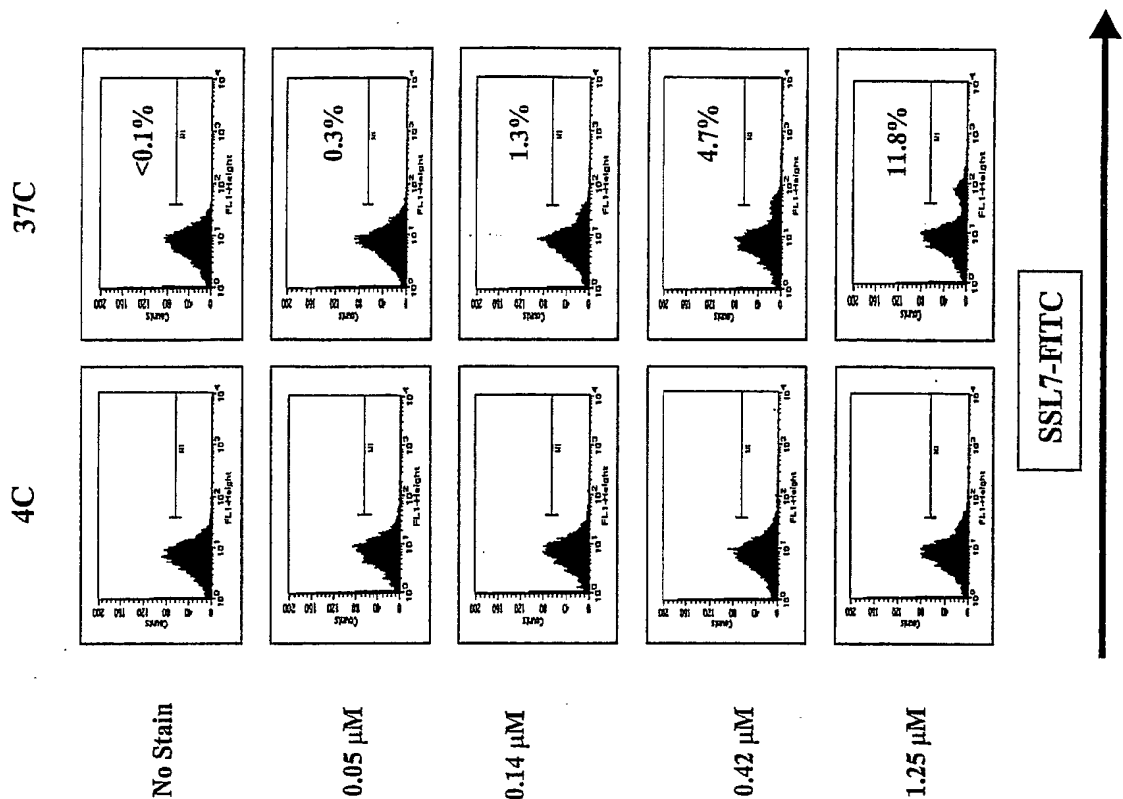


Figure 3(a)

5/14

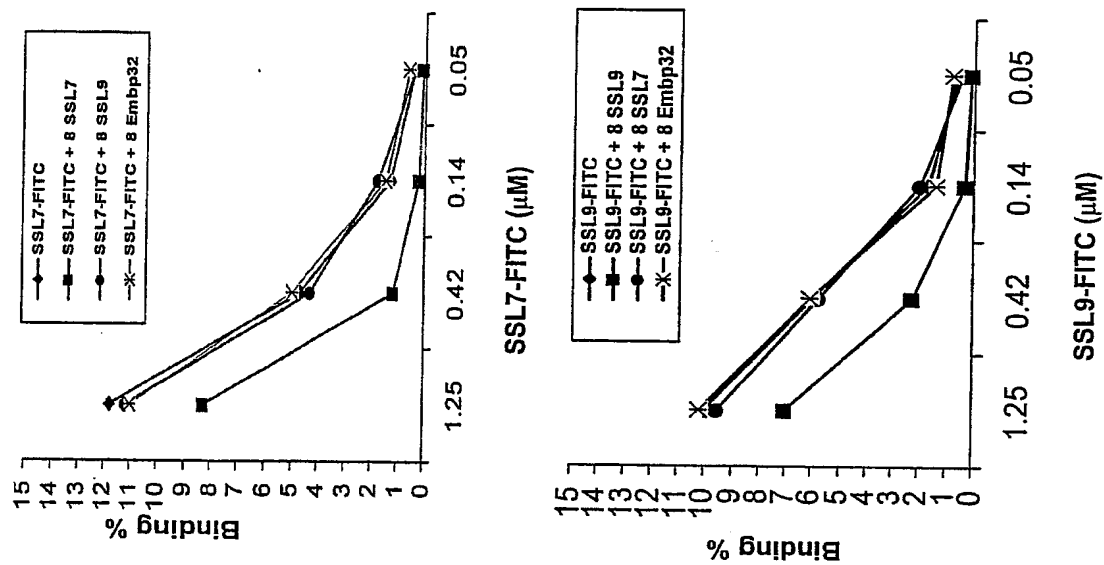


Figure 3 (b)

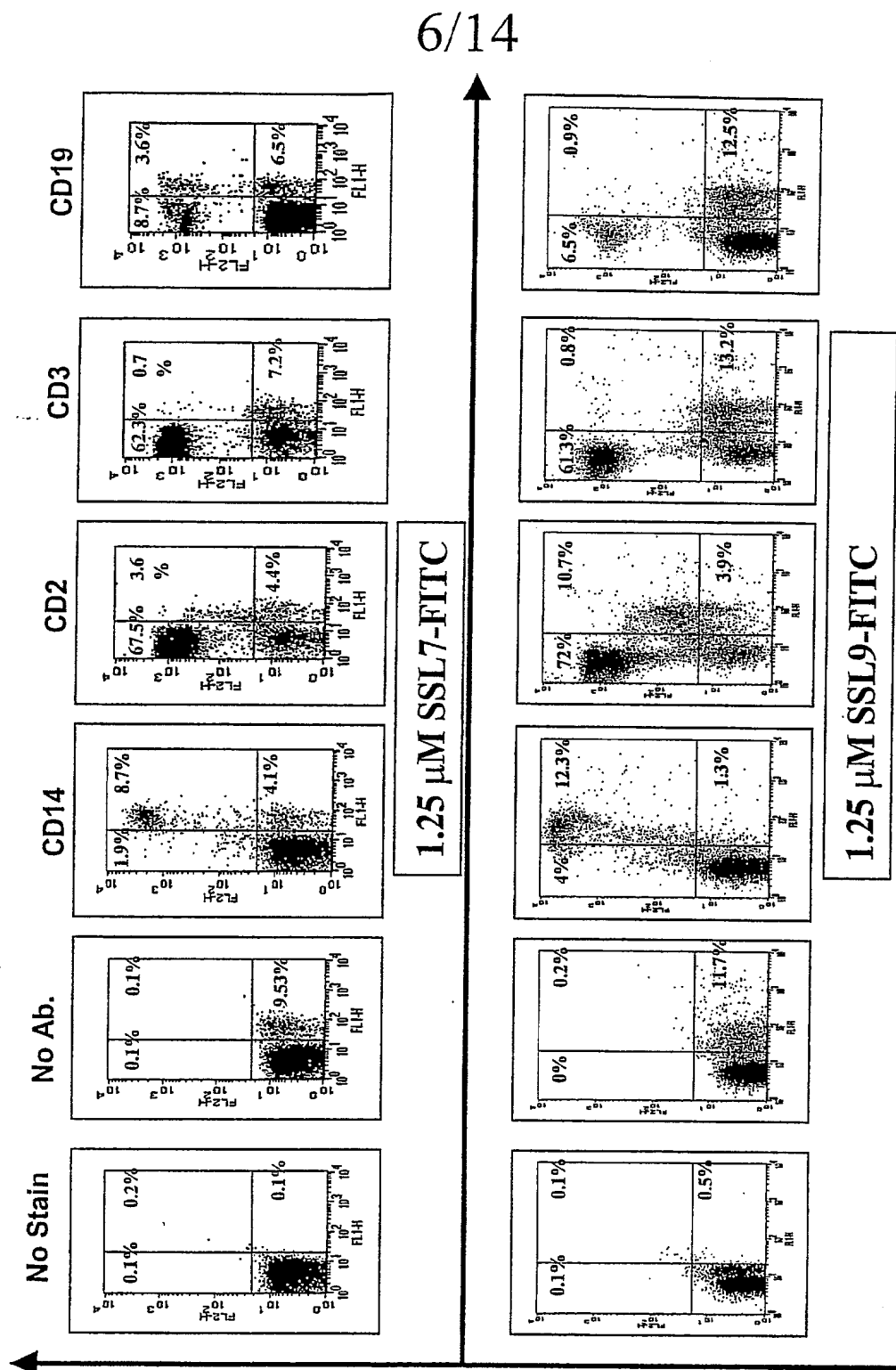


Figure 4

7/14

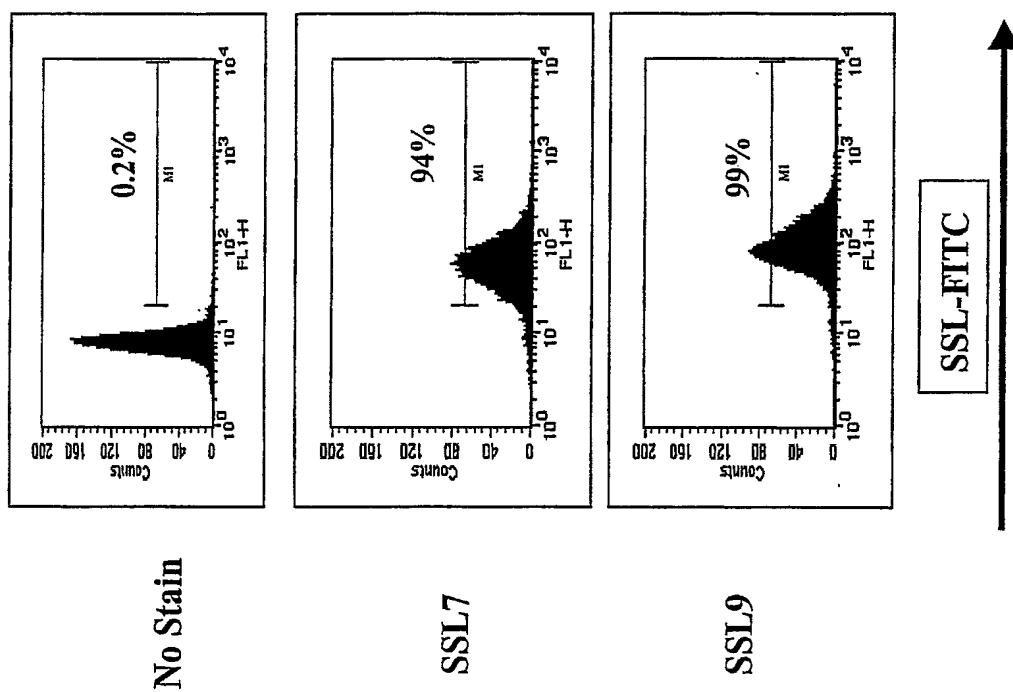


Figure 5

8/14

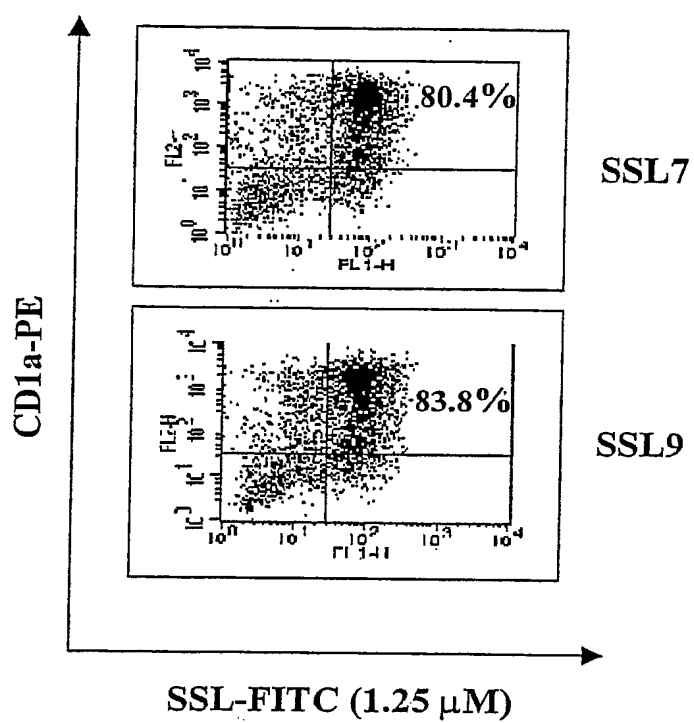
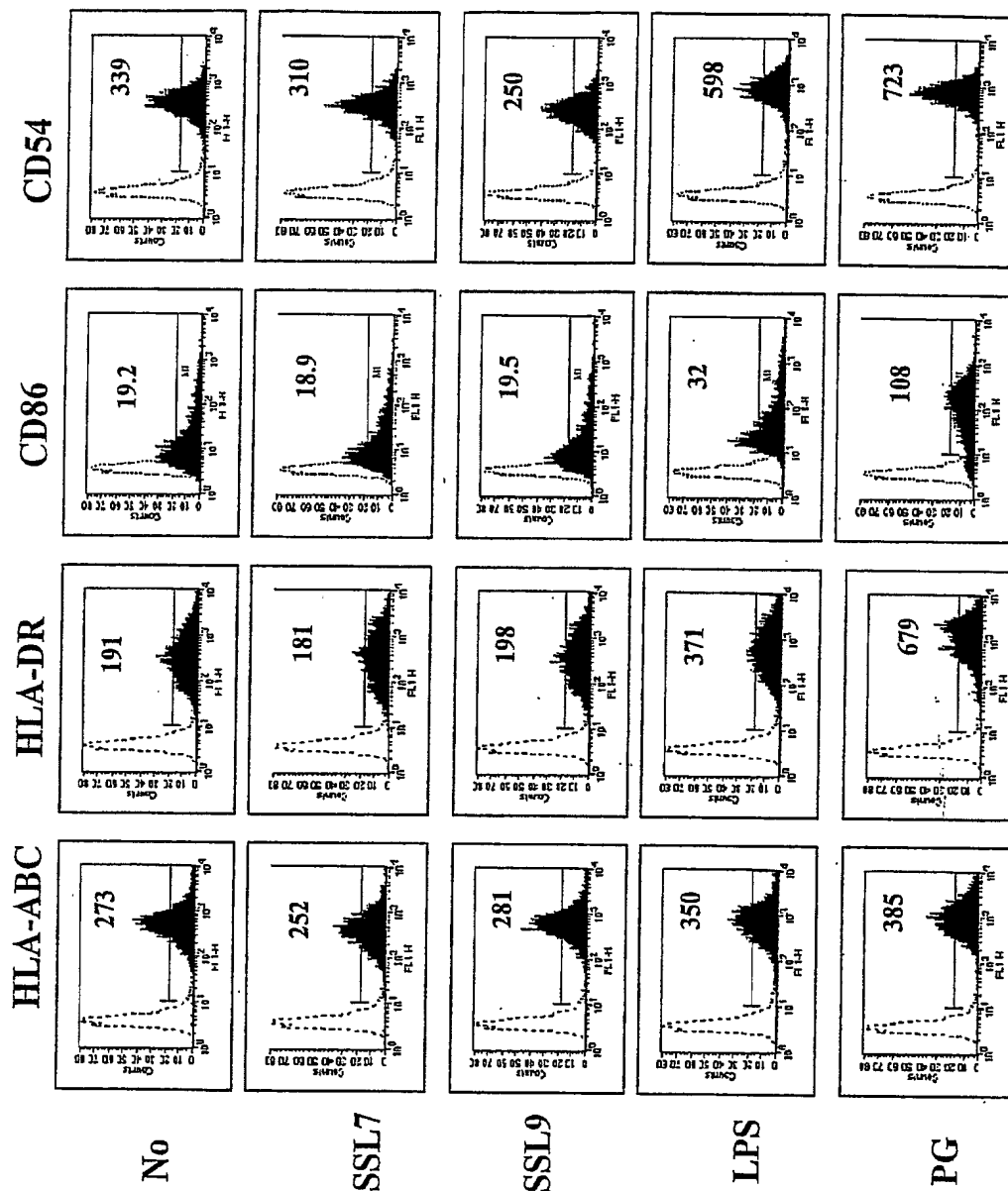


Figure 6

9/14

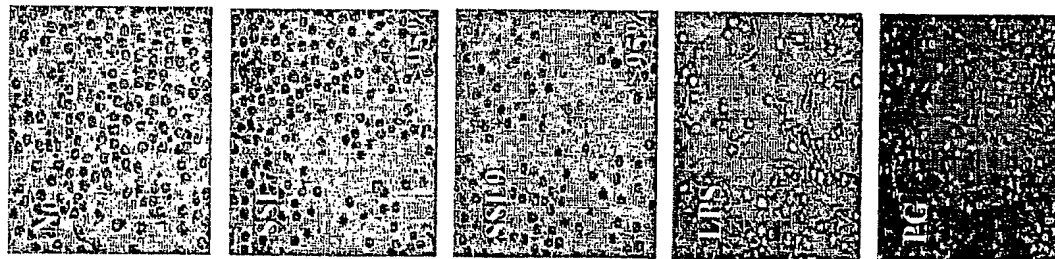
B



Median Fluorescence Intensity (MFI)

Figure 7

A



10/14

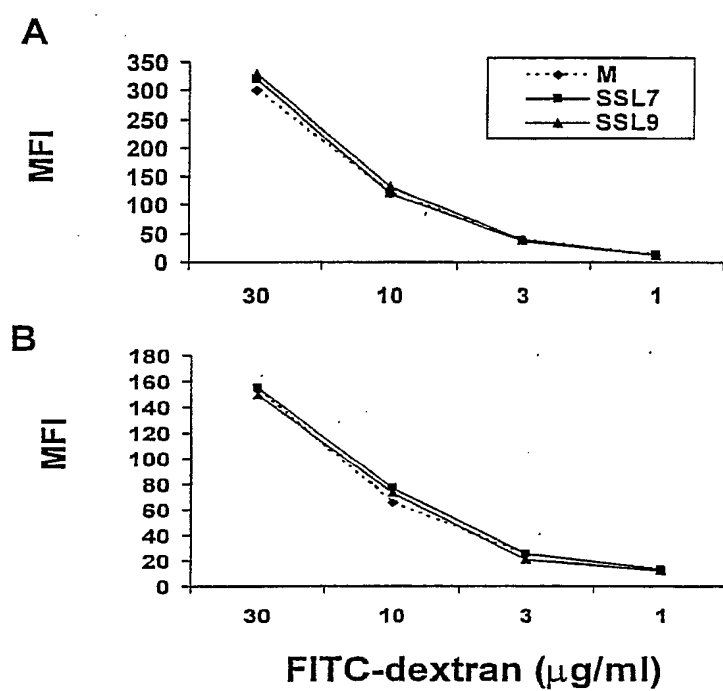


Figure 8

11/14

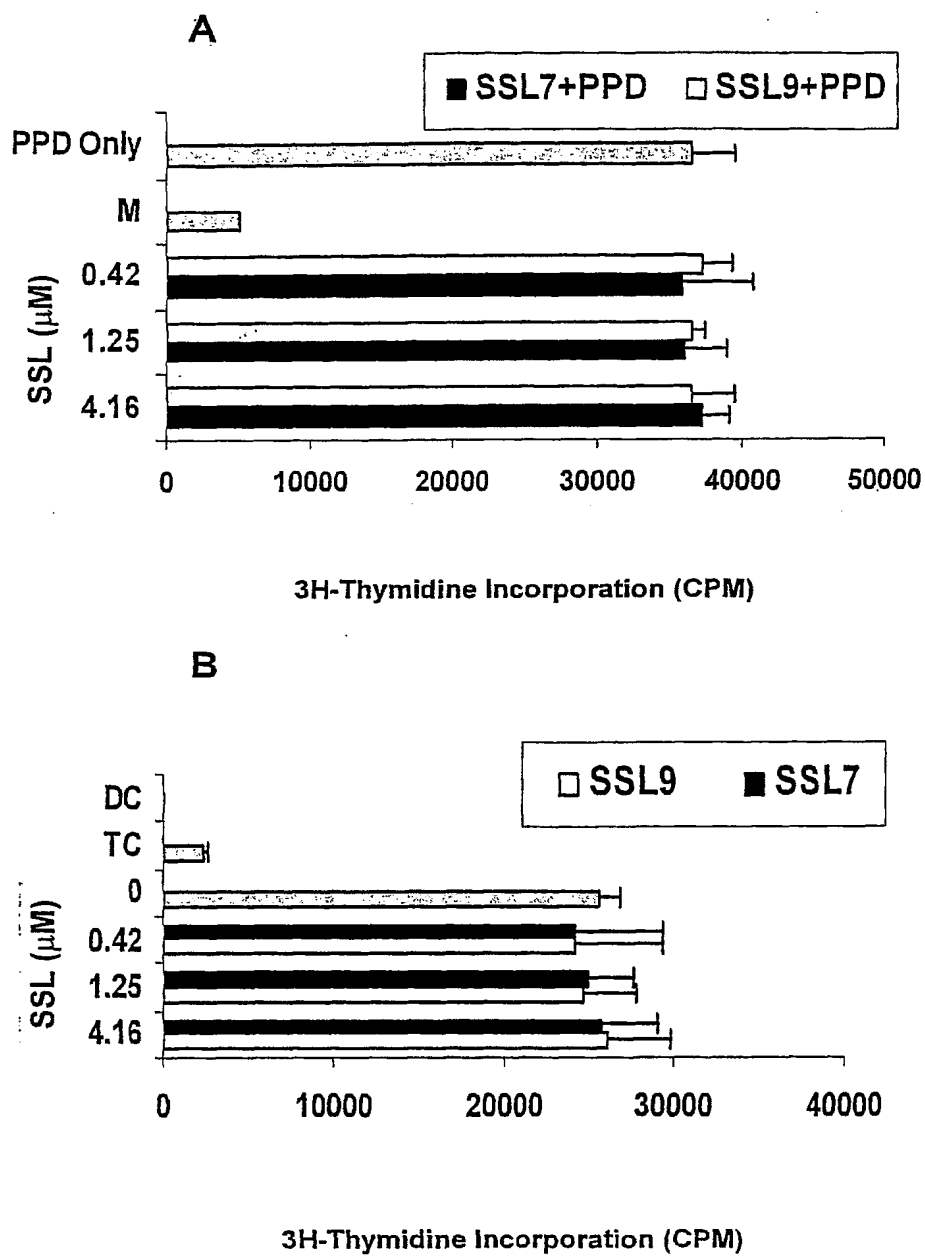


Figure 9

12/14

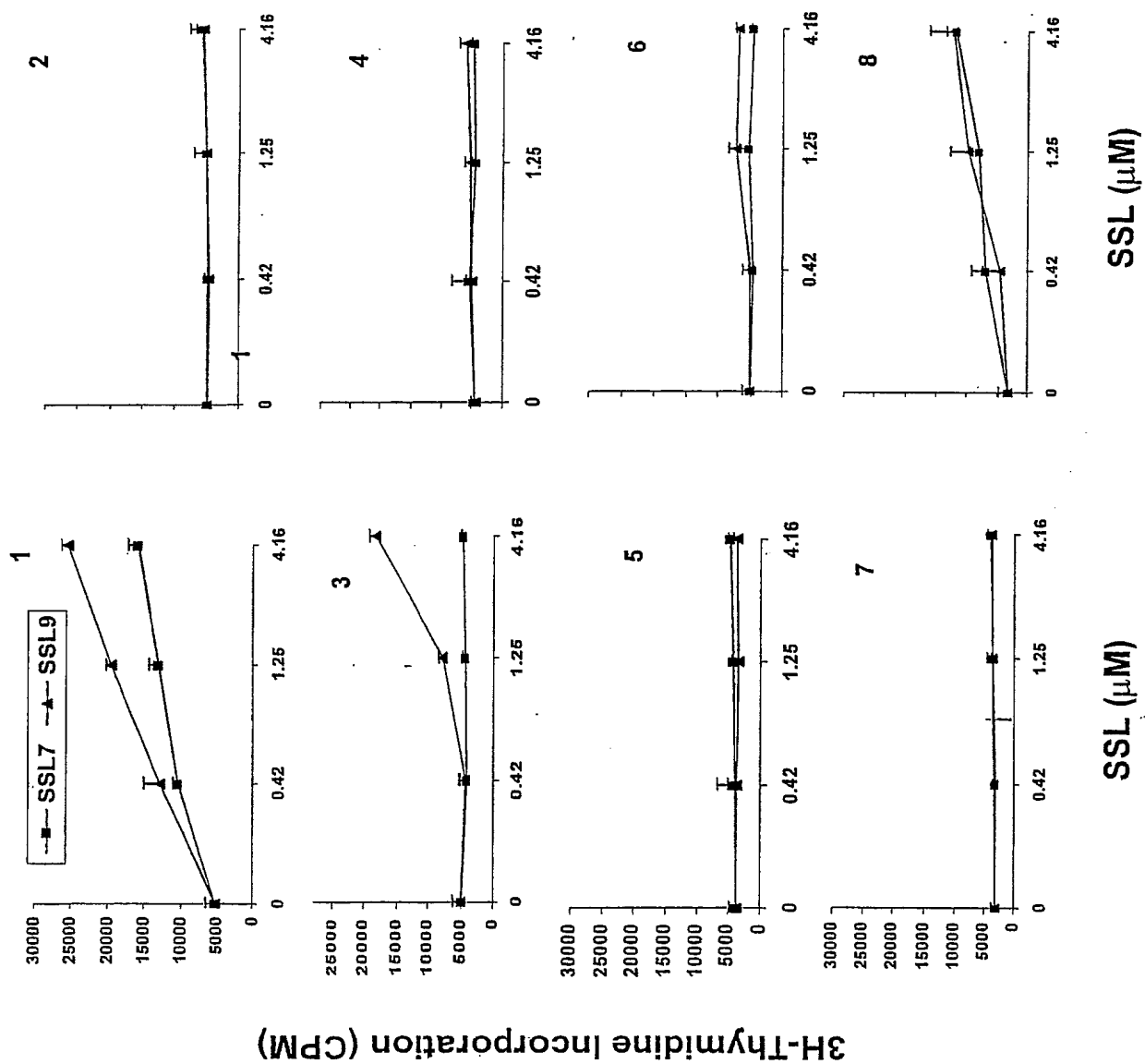


Figure 10

13/14

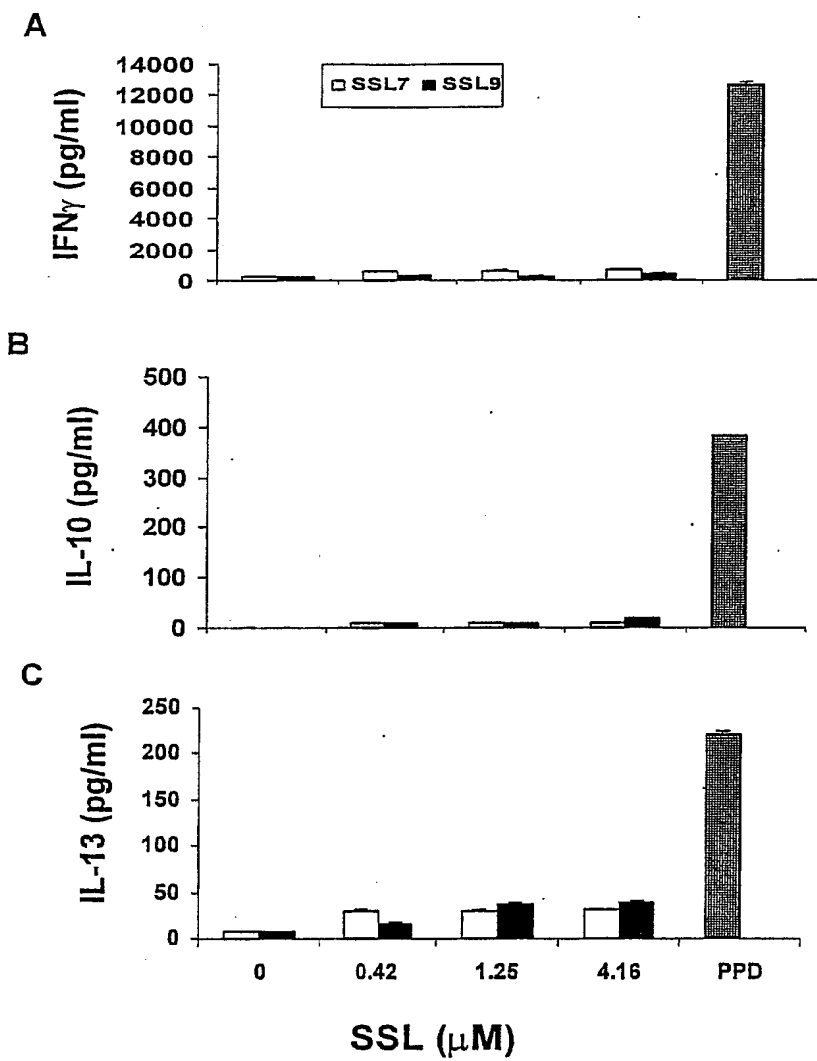
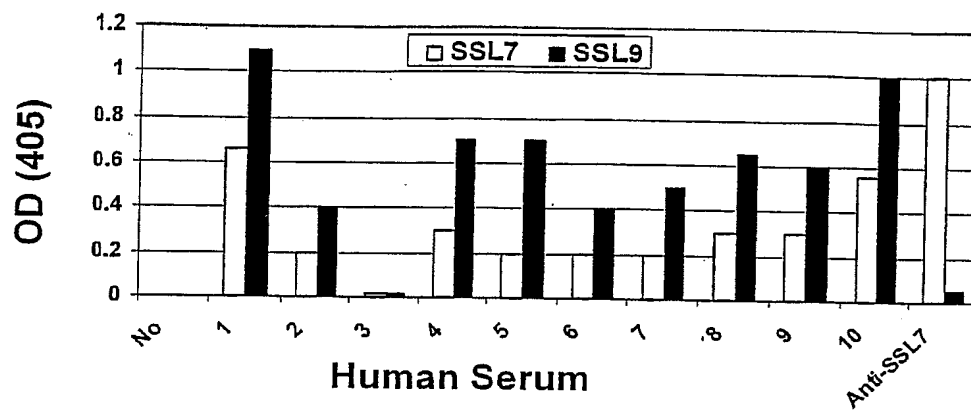


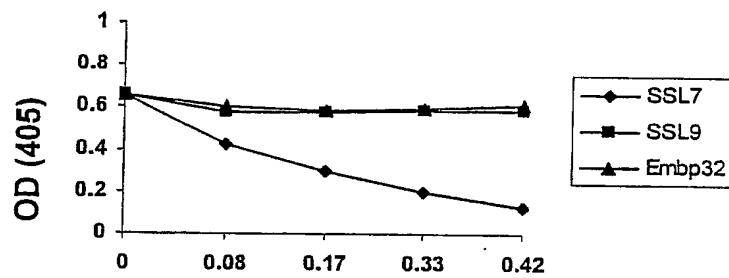
Figure 13

14/14

A



B



C

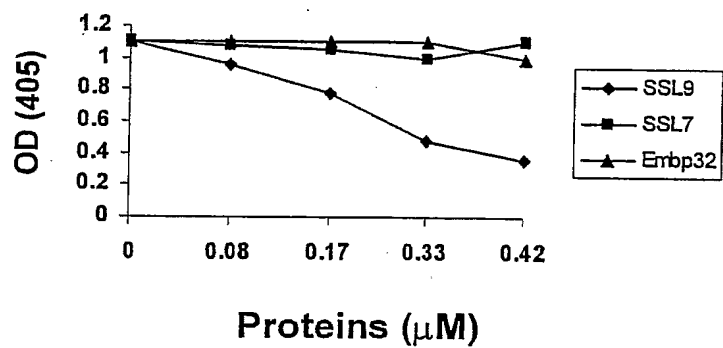


Figure 14

**SEQUENCE LISTING*****S. aureus* strain N315 taken from GenBank**

5

SEQ ID no:1

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

143101 tatgaaattt aaagcgatag caaaagcaag tttagcattg ggaatggttag caacagggtgt
143161 aattacatcg aatgtacaat cagtacaagc gaaaacagaa gttaaacaac aaagtgaatc
143221 agagttgaaa cactattata ataaaccggt tttagagcgt aaaaatgtta ctggatataa
143281 atatactgaa aaaggtaaag attatataga tgttatagta gacaatcaat attctcaaat
143341 ttcttttagtt ggatctgata aagacaaatt taaagatgga gacaactcga atatatagtg
143401 gtttatcctt agagaagggtg acagtagaca agcaacaaat tactcaattg gtggcgtaac
143461 aaaaacaaac agtcaacctt ttattgacta tatacacaca ccaatccttg aaatcaagaa
143521 aggtaaagaa gaaccacaaa gtagtttata ccaaatttat aaagaagaca tctcattgaa
143581 agaacttgat tatagattaa gagaacgtgc aatcaaacia cacggcttgt attcaaattg
143641 tcttaaacaa ggtcaaatta caattacaat gaaagatggc aaatcacata ctatcgattt
143701 aagtcacaaa cttgaaaaag aacgtatggg tgattctatc gacggcgagc aaatacaaaa
143761 aatttctagta gaaatgaaat aatactttct aacaacaaag cgctatgttg aatagtgctt
143821 gttatggaag tatatggaag ttaagcgacg tactgttgct tagcttcttt ttttgagggg
143881 aaaagttaca aaactcacac aaacagtcgc accacgcatt atcttttgct taaatagctt
143941 aatcatattt tatgaatagt taaaacaggg ttaatgtgaa tatccgaata cagctcctat
144001 aatatgggtg tatggttcaa attacgtaat aaaacaatct aattataata gattggagca
144061 tacaactatg aaaaatgaaa atattgcaaa aataagtttg ttattaggaa tattagcaac
144121 aggtgtaaac actacaacgg aaaaaccagt tcatgccgaa aagaaacctt ttgtaataag
144181 tgaaaatagc aaaaaattaa aagcttatta tactcaacct agtattgaat ataaaaatgt
144241 gacagggttat atcagtttca ttcaaccaag tattaatttt atgaatatca tagatggtaa
144301 ttctgttaat aaccttgctt taattggcaa agataagcaa cattatcata cgggtgtaca
144361 tcgtaatctt aatatatttt acgttaatga ggataagaga tttgaagggt caaagtactc
144421 tattgggggt atcactagtg caaacgataa agctgtcgac ctaatagcag aagcaagagt
144481 tattaagca gatcatattg gtgaatatga ttatgacttt ttccatttta aaatagataa
144541 agaagcgatg tcattgaaag agattgattt taaattaaga aaatacctta ttgataatta
144601 tgggtctttac ggtgaaatga gtacagggaa aattaccgtc aaaaagaaat actatggaaa
144661 gtatacattt gaattggata aaaaagtaca agaagaccgt atgtccgatg ttatcaatgt
144721 cacagatatt gatagaattg aaatcaaagt tagaaaagct taatacacat acttgacgac
144781 gaaataattt gaaattgaaa tagagagggt aagtgacgat caaacgttgc ttaacttctt
144841 tttaatgctt aaaaatcatt tcaaaggcac atagaaacgc tatattaacc tcataatcac
144901 tcattatttt ttgcttaaat tacttaataa tacttcaata attgttaaaa agggtttaat
144961 gtgattatct tagaacgcca tctataatga tgttgatga ttcaaattac gtaaaaagac
145021 aatcgaaat atatatagtt ggagtataca attatgaata tgaaaacaat tgcataaac
145081 agtttagcac tagggctttt acaacaggcg gcaattacag taacgacgca atcgggtcaaa
145141 gcagaaaaaa tacaatcaac taaagttgac aaagtaccaa cgcttaaaagc agagcgatta
145201 gcaatgataa acataacagc aggtgcaaat tcagcgacaa cacaagcagc taacacaaga
145261 caagaacgca cgcctaaatc cgaaaaggca ccaaatacta atgaggaaaa aacctcagct
145321 tccaaaatag aaaaaatata acaacctaaa caagaagagc agaaatcgct taatatatca
145381 gcaacgccag cgcctaaaca agaacaatca caaacgacaa ccgaatccac aacgccgaaa
145441 actaaagtga caacacctcc atcaacaaac acgccacaac caatgcaatc tactaaatca
145501 gacacaccac aatctccaac cataaaacia gcacaaacag atatgactcc taaatatgaa
145561 gattttaagag cgtattacac gaaacggagt tttgaatttg aaaagcagtt tggatttttg
145621 ctcaaaccat ggacgacggg taggtttatg aatgtttatc caaatagggt catctataaa
145681 atagctttag ttggaaaaga tgagaaaaaa tataaagatg gaccttacga taatatcgat
145741 gtatttatcg ttttagaaga caataaatat caattaaaaa aatattctgt cgggtggcatc
145801 acgaagacta atagtaaaaa agttgatcac aaagcagaat taagcggttac taaaaagat
145861 aatcaaggta tgatttcacg cgatgtttca gaatacatga ttactaagga agagatttcc
145921 ttgaaagagc ttgattttta attgagaaaa caacttattg aaaaacataa tctttacggt
145981 aacatggggt caggaacaat cgttattaaa atgaaaaaac gtgggaagta tacgtttgaa
146041 ttacacaaaa aactgcaaga gcatcgtagc gcagatgtca tagaagggtac aaacattgat
146101 aaaattgaag tgaatataaa ataactatga cgttctctaa atagaagctg acatcggttaa
146161 aacaagaagt taagtgacaa cggtttacat gttgcttagc ttctttttatt atgcgtaatg
146221 atgtaaaaaga cgaatattca tttgtttgta aaagtggcat ttctatgtct taaaagtgc
146281 gaatcctcaa atgtgccaag tgttgaatca catcaaaatc agtttttatt aacgaacatt
146341 atggatttct taattttactt aacgatgatt caaatatagt taaacaagggt ttaatgtgaa

```

	146401	tggagcaata	cgccatctat	aataaagctg	tatgattcaa	tgaatgtaat	cgaacaaatc
	146461	taataattac	gaatggagca	tacaactatg	aaaataacaa	caattgctaa	aacaagttaa
	146521	gcactaggcc	ttttaacaac	gggtgtaatc	acaacgacaa	cgcaagaagc	aaacgcgaca
5	146581	acaccatctt	ccactaaagt	ggaagcacca	caatcaacac	cgccctcaac	taaagtagaa
	146641	gcaccgcaat	caaaacccaa	cgcgacaaca	ccgccttcaa	ctaaagtgga	aacaccgcaa
	146701	caaacaccga	acgcgacaac	accatcttca	actaaagtgg	aaacaccaca	atcgccaacc
	146761	acaaaacaag	taccaacaga	aataaatcct	aaattttaaag	atttaagagc	gtattatagc
	146821	aaaccaagtt	tagaatttaa	aaatgagatt	ggtattatct	taaaaaaatg	gacgacaata
	146881	agattttatga	atattgttcc	agattatttc	atatataaaa	tagcttttagt	tggaaaagac
10	146941	gataaaaaat	atgatgaagg	agtacatagg	aatgtcgaatg	tatttgtcgt	tttagaagaa
	147001	aaaaataaat	atggagtggg	aagatactcg	gtcgggtgga	tcacaaagag	taatatgtaa
	147061	aaagtgtgatc	acaaagcagg	agtaagaatt	actaaagaag	ataataaagg	tacaatctct
	147121	catgatgttt	cagaattcaa	gattactaaa	gagcagattt	ccttgaaaga	acttgatttt
	147181	aaattgagaa	aacaacttat	tgaaaatcat	aatctgtacg	gtaacggttg	ttcaggtaaa
15	147241	attgttatta	acatgaaaaa	cggtggaaag	tacacgtttg	aattacacaa	aaaattacaa
	147301	gaaaaatcgca	tgccagacgt	catagatggc	actaatattg	ataacattga	agtgaatata
	147361	aaataatcat	gacattctct	aaatagaagg	tgcatcggga	aaaacaagaa	gttcagtga
	147421	aacggcctac	atgttgctta	gcttcttttg	ttatgttcga	tgatttgaga	acccgaattt
	147481	tcgatgggtc	caaatatgac	gtggaagagt	cctgaattta	tctgtaaatc	cctatctatc
20	147541	gggtgtgaag	cacaacggaa	tcagttttat	ttaacgaaca	ttatagattc	cttaatttac
	147601	ttaataatga	ttcaatgatt	attaaagatg	gtttaatgtg	aaaggtcaaa	tacgccaatt
	147661	ataataaagc	tgtatgattc	aatagacatg	agcgaacaaa	tctaataatt	acgaatggag
	147721	catacaacta	tgaaaatggc	agcaattgcg	aaagcaagtt	tagcattagg	tattttagca
	147781	acaggaacaa	taacgtcatt	gcacaaact	gtaaatgcga	gtgaacatga	agcaaaatat
25	147841	gaaaatgtga	caaaagatat	ctttgactta	agagattact	atagtgccgc	aagtaaggaa
	147901	cttaaaaaatg	ttactgggtta	tcgttatagc	aaaggtggca	agcattacct	tatctttgat
	147961	aaaaatagaa	aattcacaag	agtacagata	tttggtaaaag	atattgaaag	atttaaagca
	148021	cgcaaaaaatc	cgggattaga	catatttgtt	gttaaagaag	cagaaaatcg	caacgggcaca
30	148081	gtgttttcat	atggtggtgt	cactaaagaa	aatcaagacg	cttattatga	ttatataaac
	148141	gcaccaagat	ttcaaatcaa	gagagatgaa	ggtgacggta	ttgctacgta	cggtagagta
	148201	cactacattt	ataaagaaga	gatttcactt	aaagaactcg	actttaaatt	gagacagtat
	148261	ttaattcaaaa	attttgatct	gtataaaaag	tttcctaaaag	atagtaagat	aaaagtgata
	148321	atgaaagatg	gcggctatta	tacgtttgaa	cttaataaaa	aattacaacac	aaatcgcatg
35	148381	agtgatgtca	ttgacggtag	aaatattgaa	aaaatagaag	ccaacattag	ataattcaat
	148441	gaaatattgga	taatagtaaa	atatggatag	tatagaggag	ttaggcaaca	taagttgctt
	148501	agcttctttt	ttgtgttggc	gagatgaaaa	tgaagcgtat	cgatgaataa	taaaaaacacc
	148561	aataaaactt	gtggaatag	ttgatactta	tagatgcgtg	atgtcgtctt	agtgacatga
	148621	aaacatgtgg	aaaacataat	taaatttgagg	gaaagtgtga	atagttaaaa	aattagcatt
40	148681	gtgtttataaa	aaataattaa	tactgttagg	atttcattaa	ctaacttaac	gttggttcaa
	148741	aaatagttaa	aaagaggtta	attcatagcg	cagtatctcg	cttatataat	gatagtagat
	148801	tggttcgtatt	acgtaattga	attaatcata	taaaaaatata	ttaagacaaa	atttataaat
	148861	agattgggag	aatagtacta	tgaaattaaa	aacgttagct	aaagcaacat	tggcattagg
	148921	cttattaaact	actggtgtga	ttacatcaga	aggccaagca	gtccacgcaa	agaaaagca
45	148981	agagagagta	caacatttat	atgatattaa	agacttatat	cgatactact	catcagaaag
	149041	ttttgaattc	agtaatatga	gtggtgaagg	tgaaaactat	aacggttcta	acgttgttacg
	149101	ctttaaccaa	gaaaaacaaa	atcaccaatt	attcttatta	ggaaaagata	aagataaata
	149161	taaaaaaggc	cttgaaggcc	agaatgtctt	tgtggtaaaa	gaatttaattg	atccaaacgg
	149221	tagactatct	actggttggtg	gtgtgactaa	gaaaaataac	aaatcttctg	aaactaatat
50	149281	acattttatct	gttaataaag	tgtatggcgg	aaatttagat	gcatcaattg	actcattttt
	149341	aattaataaaa	gaagaagttt	cactgaaaga	acttgatttc	aaaattagaa	agcaatttagt
	149401	tgaaaaatat	ggtttatata	aagggtacgac	taaaatcggt	aagatcacta	tcaatttgaa
	149461	agacgagaaa	aaggaagtaa	ttgattttagg	tgataaaactg	caattcgagc	gcatgggtga
	149521	tgtgttgaa	agtaaggata	ttcaaaatat	agcagtgaat	attaatcaaa	tttaaatata
55	149581	acaatcaatg	actttaaagt	aataaatttg	aagcagctta	acgatgaaat	gttgaaataa
	149641	tacgtacatc	ctccaaaaag	ggcggtatct	aaatcaacag	tgctgttagg	ctgtttttat
	149701	gtttttataac	gcagggtatg	agcgtactaa	aaattcacat	tacttctgaa	agtgatgtcc
	149761	attgaatatt	aattagttct	tcatttaacca	tgatttaatt	tttaataaac	gagtgttaat
	149821	ctcagttctg	ctcaatgcc	tttataataa	atgtgtatta	ttcaaaattac	gtaataaaag
	149881	caatccaata	tattaagatt	ggagcatatg	aatatgaaat	ttacagcgat	agctaagcg
60	149941	atattttgat	taggaatatt	aacaacaagt	gtaatgataa	cagaaaatca	atcggttaat
	150001	gcaaaaggaa	agtatgaaaa	aatgaaccgt	ttatatgata	caacaagtt	acatcaatac
	150061	tattcaggac	ctagtattga	gttaacaaat	gttagtgcc	aaagtcaagg	ttattatgac
	150121	tctaacgttt	tgctttttta	ccaacaaaat	caaaagtctc	aagtattttt	attgggaaaa
	150181	gtgaaaaata	aatacaaga	aaaaaacacat	ggtttagatg	tctttgcggt	accagaatta
65	150241	gtagatttag	atggaagaat	atttagtgtt	agtggtgtaa	cgaagaaaaa	cgtaaaaatca

	150301	atattttgagt	ctctaagaac	gccgaactta	ctagttaaaa	aaatagacga	taaagacggt
	150361	ttttctattg	atgaattttt	ctttattcaa	aaggaagaag	tgctattgaa	ggaacttgat
	150421	tttaaaataa	gaaaactggt	gattaaaaaa	tacaaactgt	atgaagggtc	agctgataaa
	150481	ggtagaattg	ttattaatat	gaaagatgaa	aataagtatg	aaattgattt	aagtgataaa
5	150541	ttagatttcg	agcgtatggc	agatgtcatt	aatagtgaac	aaattaagaa	catcgaagtg
	150601	aatttgaaat	aatcaatgat	atatagaata	aaagcttaag	aagcggttta	ataatcccat
	150661	gtttaatgat	tttgatacgt	gttttaataa	taaaaacata	tcgaacattg	actacgttat
	150721	taagctgctt	ttttgtacac	tttgtatcga	ataacttaag	atctaaaact	aatcggaag
	150781	aacaatgatt	ccccaaaaa	aatttatggt	gctattaaaa	atcagttaat	acgaatgta
10	150841	aaatacgttt	gatttttcatt	aataatgatt	caagtttatt	taaatgagcg	ttaatgtcag
	150901	tctgttttga	tgacacctat	aataaagaca	gatagttaa	attacgtaat	cataacaatc
	150961	caatatatca	agattggagc	aaataaatat	gaaattgact	gcattagcaa	aagcaacatt
	151021	agcattagga	atattaacta	caggtgtggt	tacagcagaa	agtaaagctg	ttcacgcgaa
	151081	agtagaactt	gatgagacac	aacgcaata	ttatatcaat	atgctacatc	aatactattc
15	151141	tgaagaaagt	tttgaatcaa	caaacttag	cgttaaaagt	gaagattact	atgggtctaa
	151201	cgttttaaac	tttaaccaac	gaaataaaac	tttcaaagt	tttctgcttg	gtgacgataa
	151261	aaataaatat	aaagaaaaaa	cacatggcct	tgatgtcttt	gcagtacctg	aattaataga
	151321	tataaaagg	ggcatatata	gcgttggcgg	tataacaaag	aaaaatgtga	gatcagtggt
20	151381	tggatttgta	agtaatccaa	gtctacaagt	taaaaaagt	gatgctaaac	atggcttttc
	151441	gataaatgag	ttgttcttta	ttcaaaagga	agaagtatcg	tggaagtaac	tggattttta
	151501	aataagaaaa	atgttagtcg	aaaaatatag	attgtataaa	ggcgcgtcag	ataaaggtag
	151561	aatcgttatt	aatatgaaag	acgaaaagaa	atatgtaatt	gatttaagtg	aaaaattaag
	151621	ttttgatcgt	atgtttgatg	taatggatag	taagcaaat	aaaaatattg	aagtgaattt
	151681	gaattagttt	gagttaatag	cataatagct	taagaagcga	cttaacgaca	aaatgtgaat
25	151741	tgacatctgt	gtccttata	aaggaaactgt	gttaaatata	ttactgttgt	taagttgttt
	151801	ttgtaattca	aagagcagaa	cagagtaaca	tcactcagttg	tagtaaacga	taatccagta
	151861	aaacaactaa	atgaaataat	gaaagtcatt	taacctgaac	attaaaaat	atgtgttttt
	151921	cattaagaat	aattcaagta	tatttaaatc	gaggttaatt	atcgtatgaa	acgatgcacg
	151981	ttataataaa	aatgtatgat	tcaaattacg	taatgaaaac	aatccaat	attaagattg
30	152041	gagcaataa	atatgaaatt	tacagcatta	gcaaaagcaa	cattagcttt	aggaatttta
	152101	acaacaggaa	ctttaacaac	agaagttcat	tcaggtcatg	caaaacaaaa	tcaaaagtca
	152161	gtaaataaac	atgacaagga	ggcactatac	cgatactaca	ctggaaagac	tatggaaatg
	152221	aaaaatatta	gtgctttgaa	acatggtaaa	aataacttgc	gttttaagtt	tagagggtatt
	152281	aagattcaag	ttttactgcc	tggaaatgat	aaaagtaaat	ttcaacagcg	tagttatgag
35	152341	gggttagatg	tggtttttgt	tcaagaaaaa	agagataagc	acgatataat	ttatactggt
	152401	ggtggtgtaa	tacagaataa	taaaacatct	ggagttgtca	gtgcaccaat	attaaatttt
	152461	tcaaaagaaa	aggggtgaaga	tgcttttgtg	aaagggttacc	cttattacat	taaaaagaaa
	152521	aaaataacac	taaaagagct	ggattataag	ttgagaaagc	atctaattga	aaaatacggg
	152581	ctttataaaa	caatctcaaa	agatggtagg	gtcaaaatta	gcttgaaaga	tggcagtttt
40	152641	tataaccttg	atttaagatc	taaattaaaa	tttaaatata	tgggggaagt	catagaaagc
	152701	aaacaaatta	aagatattga	agtttaactta	aagtaaaata	ttacgaataa	taaaagataa
	152761	ttgaagcggc	tttaacgggtga	aatgtaaatt	ggtgcgcata	gottatacaa	aaaggatgca
	152821	tcaatcgata	tcgtcggtta	gccgtttttg	tttgtgtgtc	atgaatccta	tcccaatctc
	152881	catgaatata	aaaattccac	caccaacatc	aaaattctcc	acatcgcaac	ataaccaaat
45	152941	gttataataa	atctattaca	caaagagata	aattacttat	tcaaaggcgg	aggaatcaca
	153001	tgcttattac	tgaaaaacaa	cgtcagcaac	aagctgaatt	acataaaaaa	ttattggtcg
	153061	ttgcgaatga	tttaagaggg	aatatggatg	cgaagtgaatt	ccgtaattac	attttaggct
	153121	tgattttcta	tcgcttctta	tctgaaaaag	cggaacaaga	atatgcagat	gccttgtcag
	153181	gtgaagatat	tacgtatcaa	gaagcatggg	cagatgaaga	atatcgtgaa	gacttaaaaag
50	153241	ctgaattaat	tgatcaagtc	ggttacttca	ttgaaccaca	agattttattc	agtgcgatga
	153301	ttcgtgaaat	tgaaacgcaa	gatttcgata	tcgaacatct	cgcaacggcg	attcgtaaag
	153361	ttgaaacatc	aacactaggt	gaagaaaagt	aaaatgactt	tatcggaactg	ttcagcgata
	153421	tggacttaag	ttcaacgcga	ctaggtaaca	atgtcaaaaga	acgtactgca	ctaattttcca
	153481	aagttatggt	taaccttgac	gacttaccat	tcgttcacag	tgatatggaa	attgatattgt
55	153541	taggtgatgc	atacgaattt	cttatcgggc	gctttgcggc	gacagcaggt	aaaaaagcag
	153601	gcgagttcta	tacaccacaa	acagtatcta	agatactggc	gaagattgtc	acagacggta
	153661	aagataaaat	acgtcacgtg	catgacccaa	catgtgggtc	aggttcatta	ttgttacgtg
	153721	ttggtaaaga	aacgcaagt	tatcgttatt	tcgggtcaaga	acgtaacaat	accacttaca
	153781	acttagcgcg	catgaacatg	ttattacatg	atgtgcgcta	tgaaaacttt	gatataccgta
60	153841	atgatgatac	gttggaataa	ccagctttct	taggcaatac	atgtgatcg	gttatagcga
	153901	acccaccata	cagtgcgaaa	tggactgcga	attcaaaatt	tgaaaatgat	gaacgcttca
	153961	tggtgtacgg	caagcttgcg	ccaaaatcca	aagcagactt	tcgctttatt	caactattgg
	154021	tacattacct	agacgatgaa	ggtaacctgg	cagttgtact	cccacattgt	gtccttattcc
	154081	gtggtgccc	agaaggcgtc	attcgtcgtt	atttaattga	agaaaagaac	tacttagaag
65	154141	cogtgattgg	cttaccagtg	aatattttct	atgggacaag	tattccaaca	tgtatcttag

154201 tattttaaaaa atgtcgccaa caagacgaca acgtattatt tatcgaatgca tccaatgatt  
 154261 ttgaaaaagg aaaaaatcaa aaccattttaa gcgatgccc agtcgaacgt attattgaca  
 154321 catacaagcg taaagaaaca attgataaat acagttacag tgcgacatta caagagattg  
 154381 ccgataacga ttacaaccta aacataccga ggtatgtcga tacattcgaa gaagaagcgc  
 5 154441 caattgattt agatcaagtc caacaagatt tgaaaaatat cgacaaagaa atcgcagaaa  
 154501 ttgaacaaga aatcaatgca tacctgaaag aacttgggggt gttgaaagat gagtaataca  
 154561 caaaagaaaa atgtgccaga gttgagattc ccagggtttg aaggcgaatg ggaagagaag  
 154621 cagttggggg atcttacaga tagagtaatt aggaaaaata aaaacttaga atcgaaaaag  
 154681 cctttaacaa tatccggaca gttaggttta attgatcaaa cagaatattt tagtaaatca  
 10 154741 gtttcgtcga aaaatctaga aaattataca ctaataaaga atggagaatt cgcgtataac  
 154801 aaaagtattt ctaatggata cccattagggt gctattaaaa gattaactag atatgatagt  
 154861 ggtgtattgt cctctttgtt tatttgtttt tctattaaaa gtgaaatgtc taaagacttc  
 154921 atggaagcat attttgattc gacacactgg tatagagaag tttctggaat tgcagttgag  
 154981 ggtgcaagaa atcacggatt attaaatggt tctgtgaatg atttttttac tatttctaatt  
 15 155041 aaatatccaa gtttagaaga acagcaaaaa ataggcaagt tcttcagcaa actcgaccga  
 155101 caaattgaat tagaagaaca aaagcttgaa ttacttcaac aacagaaaaa aggctatatg  
 155161 cagaaaattt tctcacagga actgcgattc aaagatgaga atggtgaaga ttatccagat  
 155221 tgggaaaaata gcaaaataga aaaattatta aaagagagaa acgaacgttc tgacaaaggg  
 155281 caaatgcttt cagtaactat aaatagtggc attataaaat ttagtgaatt ggatagaaaa  
 20 155341 gataattcaa gtaaagataa aagtaattat aaagtagtta ggaaaaatga tattgcatat  
 155401 aattctatga gaatgtggca aggggctagt ggtaaatcaa attataatgg gattgttagc  
 155461 cctgcatata ctgtgcttta tccaacacaa aatactagct cattatttat tggatataag  
 155521 tttaaaacac atagaatgat tcataaattt aaaattaatt cacaaggatt aacatcagat  
 155581 acatggaact taaaatataa acaattaaaa aatataaata tagatatacc tgtattggag  
 25 155641 gaacaagaaa agataggtga tttctttaaa aaaatggata tattgataag taaacagaaa  
 155701 atgaaaattg aaatattaga aaaagagaaa caatcctttt tacaaaaaat gttcttataa  
 155761 ctttgataaa tacatagatt gcataagaat aaaatttgta taatttaaca taaaagttgt  
 155821 aaaaagtaaag tgaattaaaa acgaacatta aatttaggca ctgtgaaagc gcagtgtctt  
 155881 ttttgtgtcg aaattgtgta cagaataagt agttaataa agattaagtt gagataaagt  
 30 155941 gttattcgta aataaaagag agtagatcga taggaattga atgatattag ttaactattt  
 156001 attaaattac ttaataatga ttaattttta gttaaagtaa gtttaatgtg aagcacgacc  
 156061 attgctcatt ataataaatg aggattgttc gtattgcgta atagaataaa tcaaatagac  
 156121 taaaaaattg gagcatagaa ttatgaaatt aaaaaatatt gctaaagcaa gtttagcact  
 156181 agggattttt acaacaggta tgattacaac tactgtcag ccagtaaaag caagtacatt  
 35 156241 agaggttaga tcacaagcta ctcaagactt gagtgaatat tataaaggga gaggatttga  
 156301 attaacaatt gtaacaggat ataaatcagg aaataaagtt acatttatag ataattctca  
 156361 acaaatgat gttacattga caggaaatga aaaattaact gttaaagatg atgacgaagt  
 156421 ttctaattgt gacgtgtttg tagtaagaga aggtagtgc aaatcagcta tcacaacatc  
 156481 gattgggtgga attacaaaga caaatgggac tcaacataaa gatactgttc aaaacgttaa  
 40 156541 tttgtcagtt tctaagagta cagggtcaaca cactacttct gtgacttcag aatactatag  
 156601 catctataaa gaagaaatct cattaaaaga acttgatttt aaattaagaa agcatttaaat  
 156661 tgataaacat gatctttata agacagaacc taaagacagt aaaattagaa ttactatgaa  
 156721 aaacggtggc tattatacat ttgaattaaa taaaaaatta cagcctcatc gcatgggtga  
 45 156781 tacgattgat agtagaaata tagagaaaat tgaagtgaat ttataataat attcgaggga

## SEQ ID No:2 - SSL1

50 gene /gene="(set6) N315ssl1"  
CDS 143102..143782  
 /gene="(set6) N315ssl1"  
 /note="ORFID:SA0382  
 [Pathogenicity island SaPIn2]"  
 55 /codon\_start=1  
 /transl\_table=11  
 /product="(exotoxin 6) SSL1"  
 /protein\_id="BAB41610.1"  
 /db\_xref="GI:13700312"  
 60 /translation="MKFKAIKASLALGMLATGVITSNVQSVQAKTEVKQQSESELKH  
 YYNKPVLERKNVTGYKYTEKGKDYIDVIVDNQYSQISLVGSDKDKFKDGDNSNIDVFI  
 LREGDSRQATNYSIGGVTKTNSQPFIDYIHTPILEIKKGKEEPQSSLYQIYKEDISLK  
 ELDYRLRERAIKQHGGLYSNGLKQGQITITMKDGKSHTIDLSQKLEKERMGDSIDGRQI  
 65 QKILVEMK"

5     SEQ ID No:3 - SSL2

gene           144068..144763  
                   /gene="(set7) N315ssl2"  
       CDS           144068..144763  
                   /gene="(set7) N315ssl2"  
       10            /note="ORFID:SA0383  
                   [Pathogenicity island SaPIn2]"  
                   /codon\_start=1  
                   /transl\_table=11  
       15            /product="(exotoxin 7) SSL2"  
                   /protein\_id="BAB41611.1"  
                   /db\_xref="GI:13700313"

      /translation="MKMKNIAKISLLLGILATGVNTTTEKPVHAEKKPIVISENSKKL  
       20 KAYYTQPSIEYKNVTGYISFIQPSIKFMNIIDGNSVNNLALIGKDKQHYHTGVHRNLN  
       IFYVNEDKRFEGAKYSIGGITSANDKAVDLIAEARVIKADHIGEYDYDFFPFKIDKEA  
       MSLKEIDFKLRKYLIDNYGLYGEMSTGKITVKKKYYGKYTFELDKKLQEDRMSDVINV  
       TDIDRIEIKVRKA"

25     SEQ ID No:4 - SSL3

gene           145054..146124  
                   /gene="(set8) N315ssl3"  
       CDS           145054..146124  
       30            /gene="(set8) N315ssl3"  
                   /note="ORFID:SA0384  
                   [Pathogenicity island SaPIn2]"  
                   /codon\_start=1  
                   /transl\_table=11  
       35            /product="(exotoxin 8) SSL3"  
                   /protein\_id="BAB41612.1"  
                   /db\_xref="GI:13700314"

      /translation="MNMKTIAKTSLALGLLTTGAIIVTTQSVKAEKIQSTKVDKVPTL  
       40 KAERLAMINITAGANSATTQAANTRQERTPKLEKAPNTNEEKTSASKIEKISQPKQEE  
       QKSLNISATPAPKQEQSQTTESTTPKTKVTPPSTNTPQPMQSTKSDTPQSPTIKQA  
       QTDMTPKYEDLRAYYTKPSFEFEKQFGFLKPWTTVRFMNVI PNRFIYKIALVGKDEK  
       KYKDGPDYDNIDVFIVLEDNKYQLKKYSVGGITKTNSKKVDHKAELSVTKKDNQGMISR  
       DVSEYMITKEEISLKELDFKLRKQLIEKHNLYGNMGS GTIVIKMKNNGGKYTFELHKKL  
       45 QEHRMADVIEGTNIDKIEVNIK"

SEQ ID No:5 - SSL4

gene           146488..147366  
                   /gene="(set9) N315ssl4"  
       CDS           146488..147366  
                   /gene="(set9) N315ssl4"  
                   /note="ORFID:SA0385  
                   [Pathogenicity island SaPIn2]"  
       55            /codon\_start=1  
                   /transl\_table=11  
                   /product="(exotoxin 9) SSL4"  
                   /protein\_id="BAB41613.1"  
                   /db\_xref="GI:13700315"

      60            /translation="MKITTIKTSALALGLLTTGVITTTTQEANATTPSSTKVEAPQST  
       PPSTKVEAPQSKPNATTPPSTKVETPQQTPNATTPSSTKVETPQSPTTKQVPTEINPK  
       FKDLRAYYTKPSLEFKNEIGIILKKWTTIRFMNIVPDYFIYKIALVGKDDKKYDEGVH  
       RNVDFVVL EEKNKYGV ERYSVGGITKSNKKVDHKAGVRITKEDNKG TI SHDVSEFK  
       ITKEQISLKELDFKLRKQLIENHNLYGNVSGSKIVINMKNNGGKYTFELHKKLQENRMA

DVIDGTNIDNIEVNIK"

5 SEQ ID No:6 - SSL5

gene 147730..148434  
 /gene="set10 N315ssl5"  
CDS 147730..148434  
 /gene="set10 N315ssl5"  
 /note="ORFID:SA0386  
 [Pathogenicity island SaPIIn2]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="(exotoxin 10)SSL5"  
 /protein\_id="BAB41614.1"  
 /db\_xref="GI:13700316"  
 /translation="MKMAAIAKASLALGILATGTITSLHQTVDNASEHEAKYENVTKDI  
 FDLRDYSGASKELKNVTGYRYSKGGKHYLIFDKNRKFTRVQIFGKDIERFKARKNPG  
 LDIFVVKEAENRNGTVFSYGGVTKKNQDAYDYINAPRFQIKRDEGDGIATYGRVHYI  
 YKEEISLKELDFKLRQYLIQNFIDLYKKFPKDSKIKVIMKDGYYTFELNKKLQTNRMS  
 DVIDGRNIEKIEANIR"

25 SEQ ID No:7 - SSL7

gene 148880..149575  
 /gene="set11 N315ssl7"  
CDS 148880..149575  
 /gene="set11 N315ssl7"  
 /note="ORFID:SA0387  
 [Pathogenicity island SaPIIn2]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="(exotoxin 11)SSL7"  
 /protein\_id="BAB41615.1"  
 /db\_xref="GI:13700317"  
 /translation="MKLKTAKATLALGLLTGTGVTSEGOAVHAKKEQERVQHLYDIK  
 DLYRYYSSSEFESNISGKVENYNGSNVVRFNQEKQNHQLFLLGKDKDKYKKGLEGON  
 VFVVKELIDPNGLSTVGGVTCKNNKSSETNTHLFVNKVYGGNLDASIDSFLINKEEV  
 SLKELDFKIRKQLVEKYGLYKGTTKYKGTITNLKDEKKEVIDLGDQLQFERMGDVLS  
 KDIQNIQAVTINQI"

45 SEQ ID No:8 - SSL8

gene 149914..150612  
 /gene="set12 N315ssl8"  
CDS 149914..150612  
 /gene="set12 N315ssl8"  
 /note="ORFID:SA0388  
 [Pathogenicity island SaPIIn2]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="(exotoxin 12) SSL8"  
 /protein\_id="BAB41616.1"  
 /db\_xref="GI:13700318"  
 /translation="MKFTAIAKAIFVLGILTTSVMITENQSVNAKGKYKMNRLYDTN  
 KLHQYYSGPSYELTNVSGSQSGYYDSNVLLFNQONQKFQVFLGKDKENKYKEKTHGLD  
 VFAVPELVLDGRIFSVSGVTCKNVKSIFESLRTPNLLVKKIDDKDGFSIDEFFFIQK  
 EEVSLKELDFKIRKLLIKKYKLYEGSADKGRIVINMKDENKYEIDLSDKLDFERMADV  
 INSEQIKNIEVNLK"

65

SEQ ID No:9 - SSL9

5            gene            150989..151687  
              /gene="set13 N315ss19"  
              CDS            150989..151687  
              /gene="set13 N315ss19"  
              /note="ORFID:SA0389  
              [Pathogenicity island SaPIIn2]"  
              /codon\_start=1  
 10           /transl\_table=11  
              /product="(exotoxin 13)SSL9"  
              /protein\_id="BAB41617.1"  
              /db\_xref="GI:13700319"  
 15        /translation="MKLTALAKATLALGILTTGVFTAESKAVHAKVELDETQPKYYIN  
              MLHQYYSEESFESTNISVKSEDDYGSNVLFNQNRNKTfKVFLLGDDKNKYKEKTHGLD  
              VFAVPELIDIKGGIYSVGGITKKNVRSVFGFVSNPSLQVKKVDAKHGFSINELFFIQK  
              EEVSLKELDFKIRKMLVEKYRLYKYGASDKGRIVINMKDEKKYVIDLSEKLSFDRMFDV  
              MDSKQIKNIEVNLN"  
 20

SEQ ID No:10 - SSL10

25           gene            152053..152736  
              /gene="set14 N315ss110"  
              CDS            152053..152736  
              /gene="set14 N315ss110"  
              /note="ORFID:SA0390  
              [Pathogenicity island SaPIIn2]"  
              /codon\_start=1  
 30           /transl\_table=11  
              /product="(exotoxin 14)SSL10"  
              /protein\_id="BAB41618.1"  
 35           /db\_xref="GI:13700320"  
              /translation="MKFTALAKATLALGILTTGTLTTEVHSGHAKQNQKSVNKHDK  
              LYRYYTGKTMEMKNISALKHKGKNNLRFKFRGIKIQVLLPGNDKSKFQORSYEGLDVFF  
              VQEKRDKHDIFYTVGGVIONNKTSGVVSAPILNISKEKGEDAFVKGPYYIKKEKITL  
 40           KELDYKLRKHLIEKYGLYKTISKDGRVKISLKDGSFYNLDLRSKLKFKYMGEVIESKQ  
              IKDIEVNLK"

SEQ ID No:11 - SSL11

45           gene            156143..156826  
              /gene="(set15) N315ss111"  
              CDS            156143..156826  
              /gene="(set15) N315ss111"  
              /note="ORFID:SA0393  
              [Pathogenicity island SaPIIn2]"  
              /codon\_start=1  
 50           /transl\_table=11  
              /product="(exotoxin 15)SSL11"  
              /protein\_id="BAB41622.1"  
 55           /db\_xref="GI:13700324"  
              /translation="MKLKNIKASLALGILTTGMITTTAQPVKASTLEVRSQATQDLS  
              EYYKGRGFELTNVTGYKYGNKVTFIDNSQQIDVTLTGNEKLTVKDDDEVSNVDVFVVR  
              EGSDKSAITTSIGGITKTNGTQHKDTVQNVNLSVSKSTGQHTTSVTSEYYSIYKEEIS  
 60           LKELDFKLKRLHLDKHDLYKTEPKDSKIRITMKNGGYYTFELNKKLQPHRMGDTIDSR  
              NIEKIEVNL"

SEQ ID No:12

65           244981 agtgcacatgaa gtataagtca ccttcacata ctaatacaag aggacgtcaa cagttatattt

245041 attaggattt ttaacataaa catttgctag atctgaatgt aatcttttgc ttaaatcaat  
 245101 agtgtagtta ttaccgccac cggatgatctt aagcttacct ttattacgat tticgggtata  
 245161 taatatttta ttttttatta acgcttctcg tgcacggaaa tcgatttctt tcaatgttaa  
 245221 tactggttta ttgccttggg atattttatg cgcaccaata atcgtttgta gtttatcttt  
 245281 gtattgcaca aaaagattat aagttttatc agaaggtttt gcggctgggt taacgccacc  
 245341 tgtaaatgtc tctctataag accaccataa ctgacagta tctttgtctt ttagtccaaa  
 245401 cacatctacg taacgatctt ttaactgatt aatatttccc caactttcag cgccccataa  
 245461 agatataatg gctgaccatg aatatctctt aagttcaaca taaatgtttc cattatcata  
 245521 ttgatatagc cattttattg aaaatgaaaa atgaggctgt gtgtaattat taattaattc  
 245581 attgatgtta gtttcatctt gaccaatgct ataagcttta gcttcagagt aaaaactaaa  
 245641 tactggtttt tgattttgag gtaatacagt acctaatagt aataatgttg tcgttaaaat  
 245701 tatatttttc gtgatgttct tactcattag aacatctcct ttcagaggaa tcatgatagc  
 245761 aggaataaga aattttaaag tgagcggaag caatatagta tttgcgatta tttttattaa  
 245821 ctccattatt attgttagtt tgatttttcg aggataactt caatttttgc attttgagggt  
 245881 ttttttaacat atctatttgc atcagttgat ggcaaccttt tacttaaatc tattgtgtag  
 245941 ttattgtctg tacctgttat ttttaattgt cttttattat aagaattatt atataaattt  
 246001 ttacttttaa ttaatgtttg acgaatacga aaatctaatt cttttaaagt taaaacaggc  
 246061 ttattgcctt cataaactgg aaatcgcga gtaaacgttt ctgctttatc tttatatgtt  
 246121 acattcagtt tatagtgttt atcgttagat gttgctgcag gagtaacacc accagttaac  
 246181 gtttcttgag ataatgcaaa agaatacagt gtttcttggg cttttatgcc aaaaatatca  
 246241 acgcttttat ttcttaattg gttgatattg ccccaacttt cagggtccata aacttgaata  
 246301 tgactatacc aagaaaactg taacaacgtt gcatgaatcg taccgttatc tttttgccat  
 246361 aacgtactgt tagagaaggt taaatatttt tgcgagtaat atttagttaa ctcattaacg  
 246421 tttagtttcgt tttgatttat ataataagct ttcgcttcag atgaagaatt gataggtgta  
 246481 ttaggaaaat gtgtaaatgc tgtacctaat agtaacaatg ttgttgataa aataattttt  
 246541 ttcgtgatgt tattgttcat tcgaattttc cctttgagta ttgttggaat ttttaattat  
 246601 aaaaaggggt taattagata attgaaatta tccgcattta caaaaaggtaa taggttagtt  
 246661 agatttttgc agtatgactt caatttgtgc attttttaga tttttaacat aacggtttgt  
 246721 gtcagttaat ttttaacttt tacttaaatc aatcgtgtaa ttatttccat cagcagttat  
 246781 cttaatttga cctttattaa attctccgtt atataacttt ttattcttta ttaatgtttg  
 246841 acgaatacgg aaatctaatt cttttaaagt taatactggt ttatttccct tgtaaaattc  
 246901 atgtccaccg atgatagttt gttgtttatc actatatttt aaaaatagtc tataagggtt  
 246961 atcagatgaa gtagctgctg gcgtaacacc accagtaaat gtttcatcat aagtccagta  
 247021 accttcaact gtgtcctcat ctttagttcc aaatatatca acgtatttat ttcttaactg  
 247081 attaatgttt ccccaactct cggatccaaa cacttgaata tgactatacc aaaccacagt  
 247141 ttgcaatggt gcatgaatgc taccattggg cttttgccat agccattttc cagataatga  
 247201 aaaaatcgccg tgagtgtaat atttgattaa ctcatgtgata ttagtctcgt tttcactgat  
 247261 attataggct tttgcttcag atgaaaaact gattgggtgt tttaggaagt gtgttgatgt  
 247321 ggttcctaaa agtaacaatg ctgttgataa aactaattta ttcattgatgt tctttttcat  
 247381 atgaaaaatc cctttgcgtg aattacccaa agtatataag ctattacacc gattcggaa  
 247441 taaataaaag ctaaaactat gttaaataaa cttaaacagt tagtagtgtt atttaagcaa  
 247501 aacttatcat ttttaagttt tggacagaaa cagtacttaa taaagtaggc gggagtata  
 247561 tattaataaac gacacgttac aattattctc taatcaattg cattaaattg tttgataatt  
 247621 gaattttcta actacttgaa aaatagttat actttaaatg tagtacttat ttttaattat  
 247681 tcctactact taaatttaatt attaaataaa tgttcattta attattgata aaatattaca  
 247741 aattttaata gtaggttgtg tttattttgt atgcgcttac aatttaggtg taactaaaat  
 247801 aaaaaggagt gttattaatg aaaaatttac gaaacagaag ttttttaact ttattagact  
 247861 tttcacgaca agaggtagaa ttcttattaa cactctccga ggatttaaaa cgtgctaaat  
 247921 atattggcac tgaaaagcct atgttaaaaa ataaaaatat tgcaactgta tttgaaaaag  
 247981 attctacaag aacgcgatgt gcatttgaag ttgcagcgca tgatcaaggt gcaaatgtaa  
 248041 cttattttag cccaactgga tcacaaatgg gtaaaaaaga aacaactaaa gatactgcac  
 248101 gtgtgcttgg tggaatgtat gatggcattg aataccgtgg tttttcacia agaacagtag

55 SEQ ID No:13 - SSL14

gene complement (246655..247380)  
 /gene="SA1011 N315ssl14"  
 CDS complement (246655..247380)  
 /gene="SA1011 N315ssl14"  
 /note="ORFID:SA1011"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB42263.1"  
 /db\_xref="GI:13700967"  
 /product="(SA1011) SSL14

/translation="MKKNIMNKLVLSTALLLLGTTSTQLPKTPISFSSEAKAYNISEN  
 ETNINELIKYYTQPHFSLSGKWLWQKPNQSIHATLQTVWVYSHIQVFGSESWGNINQL  
 RNKYVDIFGTKDEDTVEGYWYDETFTGGVTPAATSSDKPYRLFLKYSQKQQTIIIGGH  
 70 EFYKGNKPVLTLLKELDFRIQTLLKNNKLYNGEFNKGQIKITADGNNTIDLSKKLKL  
 TDTNRYVKNPKNAQIEVILEKSN"

SEQ ID No:14 - SSL13

5        gene                complement (245835..246560)  
          CDS                /gene="(SA1010) N315ss113"  
                              complement (245835..246560)  
                              /gene="(SA1010) N315ss113"  
                              /note="ORFID:SA1010"  
                              /codon\_start=1  
 10                            /transl\_table=11  
                              /protein\_id="BAB42262.1"  
                              /db\_xref="GI:13700966"  
                              /product="(SA1010) SSL13"

15        /translation="MNNNITKKIILSTTLLLLGTAFTQFPNTPINSSSEAKAYYINQN  
 ETNVNELTKYYSQKYLTFNSNLTWQKDNGTIHATLLQFSWYSHIQVYGPESWGNNINQL  
 RNKSVDFIGIKDQETIDSFALSQETFTGGVTPAATSNDKHYKLNVTYKDKAETFTGGF  
 PVYEGNKPVLTLLKELDFRIRQTLIKSKKLYNNSYNKGQIKITGTDNNYTIDLSKRLPS  
 TDANRYVKKPQNAKIEVILEKSN"

SEQ ID No:15 - SSL12

gene                complement (245011..245727)  
 25        CDS                /gene="(SA1009)N315ss112"  
                              complement (245011..245727)  
                              /gene="(SA1009)N315ss112"  
                              /note="ORFID:SA1009"  
                              /codon\_start=1  
 30                            /transl\_table=11  
                              /protein\_id="BAB42261.1"  
                              /db\_xref="GI:13700965"  
                              /product="(SA1009) SSL12"

35        /translation="MSKNITKNIILTTTLLLLGTVLPQNQKPVFSFYSEAKAYSIGQD  
 ETNINELIKYYTQPHFSFSNKWLYQYDNGNIYVELKRYSWSAHISLWGAESWGNINQL  
 KDRYVDVFGCLKDKDQLWWSYRETFTGGVTPAAKPSDKTYNLFVQYKDKLQTIIGAH  
 KIYQGNKPVLTLLKEIDFRAREALIKNKILYTENRNKGKCLKITGGGNNYTIDLSKRLHS  
 DLANVYVKNPNKITVDVLF"

**S. aureus strain Mu50 taken from GenBank**SEQ ID No:16

45        116881 aaaaatcaaa tttaaataga ttggggctaa aaattatgaa atttaaagcg atagcaaaag  
          116941 caagttagc attgggaatg ttagcaacag gtgtaattac atcgaatgta caatcagtac  
          117001 aagcgaaaac agaagttaaa caacaaagtg aatcagagtt gaaacactat tataataaac  
          117061 cggttttaga gcgtaaaaat gttactggat ataaatatac tgaaaaaggt aaagattata  
          117121 tagatgttat agtagacaat caatattctc aaatttcttt agttggatct gataaagaca  
 50        117181 aatttaaaga tggagacaac tcgaatatag atgtgtttat ccttagagaa ggtgacagta  
          117241 gacaagcaac aaattactca attgggtggc taacaaaaac aaacagtcaa ccttttattg  
          117301 actatataca cacaccaatc cttgaaatca agaaaggtaa agaagaacca caaagtagtt  
          117361 tataccaaat ttataaagaa gacatctcat tgaaagaact tgattataga ttaagagaac  
          117421 gtgcaatcaa acaacacggc ttgtattcaa atggtcttaa acaaggtcaa attacaatta  
          117481 caatgaaaga tggcaaatca cataactatc attttaagtca aaaacttgaa aaagaacgta  
 55        117541 tgggtgattc tatcgacggc agacaaatc aaaaaattct agtagaaatg aaataatact  
          117601 ttctaacaac aaagcgctat gttgaatagt gcttggtatg gaaatatatg gaagttaagc  
          117661 gacgtactgt tgcttagctt ctttttttga ggggaaaagt tacaaaactc acacaaacag  
          117721 tcgcaccacg cattatcttt tgcttaataa gcttaatcat attttatgaa tagttaaaaa  
          117781 cagggttaatg tgaatatccg aatacagctc ctataatatg ggtgtatggt tcaaattacg  
 60        117841 taataaaaaca atctaattat aatagattgg agcatacaac tatgaaaatg aaaaatattg  
          117901 caaaaataag tttgttatta ggaattatag caacagggtg aaacactaca acggaaaaac  
          117961 cagttcatgc cgaaaagaaa cctattgtaa taagtgaaaa tagcaaaaaa ttaaaagctt  
          118021 attatactca acctagtatt gaatataaaa atgtgacagg ttatatcagt ttcattcaac  
          118081 caagtattaa atttatgaat atcatagatg gtaattctgt taataacctt gctttaattg

	118141	gcaaagataa	gcaacattat	catacgggtg	tacatcgtaa	tcttaatatata	ttttacgtta
	118201	atgaggataa	gagatttgaa	ggtgcaaagt	actctattgg	gggtatcact	agtgcacacg
	118261	ataaagctgt	cgaccttaata	gcagaagcaa	gagttattaa	agcagatcat	attgggtgaat
5	118321	atgatttatga	ctttttccca	tttaaaatag	ttaaagaagc	gatgtcattg	aaagagattg
	118381	attttaaatt	aagaaaatac	cttattgata	attatggtct	ttacggtgaa	atgagtacag
	118441	ggaaaattac	cgtcaaaaag	aaatactatg	gaaagtatac	atttgaattg	gataaaaagt
	118501	tacaagaaga	ccgtatgtcc	gatgttatca	atgtcacaga	tattgtataga	attgaaatca
	118561	aagttagaaa	agcttaatac	acatacttga	cgacgaaata	atttgaaatt	gaaatagaga
10	118621	ggttaagtga	cgatcaaacg	ttgcttaact	tctttttaat	gcttaaaaaat	catttcaaag
	118681	gcacatagaa	acgctatatt	aacctcataa	tcaactcatta	ttttttgctt	aaattactta
	118741	ataatacttc	aataattggt	aaaaaggggt	taatgtgatt	atcttagaac	gccatctata
	118801	atgatgttgt	atgattcaaa	ttacgtaaaa	agacaatcga	atataatata	gattgtgagta
	118861	tacaattatg	aatatgaaaa	caattgctaa	aaccagttta	gcactagggc	ttttaacaac
15	118921	aggcgcaatt	acagtaacga	cgcaatcggg	caaagcagaa	aaaatacaat	caactaaagt
	118981	tgacaaagta	ccaacgctta	aagcagagcg	attagcaatg	ataaacataa	cagcaggtgc
	119041	aaattcagcg	acaacacaag	cagctaacac	aagacaagaa	cgcacgccta	aactcgaaaa
	119101	ggcaccacaa	actaatgagg	aaaaaacctc	agcttccaaa	atagaaaaaa	tatcaccaacc
	119161	taaacagaag	gagcagaaat	cgcttaatat	atcagcaacg	ccagcgccta	aacaagaaca
20	119221	atcacaaacg	acaaccgaat	ccacaacgcc	gaaaactaaa	gtgacaacac	ctccatcaac
	119281	aaacacgcca	caaccaatgc	aatctactaa	atcagacaca	ccacaatctc	caaccataaa
	119341	acaagcacaa	acagatatga	ctcctaaata	tgaagattta	agagcgtatt	acacgaaacc
	119401	gagttttgaa	tttgaaaaagc	agtttggatt	tttgctcaaa	ccatggacga	cggttaggtt
	119461	tatgaatggt	attccaaaata	ggttctacta	taaaatagct	ttagtgtgaa	aagatgagaa
25	119521	aaaaataaaa	gatggacctt	acgataatat	cgatgtattt	atcgttttag	aagacaataa
	119581	atatcaatta	aaaaaatatt	ctgtcgggtg	catcacgaag	actaatagta	aaaaagttga
	119641	tcacaaagca	gaattaagcg	ttactaaaaa	agataatcaa	ggtatgattt	cacgcgatgt
	119701	ttcagaatac	atgattacta	aggaagagat	ttccttgaaa	gagcttgatt	ttaaattgag
	119761	aaaaacaactt	attgaaaaac	ataatcttta	cggtaacatg	ggttcaggaa	caatcgttat
30	119821	taaaatgaaa	aacgggtggga	agtatacgtt	tgaattacac	aaaaaactgc	aagagcatcg
	119881	tatggcagat	gtcatagaag	gtacaaacat	tgataaaaatt	gaagtgaata	taaaataatc
	119941	atgacgttct	ctaaatagaa	gctgacatcg	gtaaaaacaag	aagttaagtg	acaacggttt
	120001	acatgtttgct	tagcttcttt	tattattcgt	aatgatgtaa	aagacgaata	ttcatttggt
	120061	tgtaaaagtg	gcatttctat	gtcttaaaag	tgacgaatcc	tcaaatgtgc	caaggttgtga
35	120121	atcacatcaa	aatcagtttt	atttaacgaa	cattatggat	ttcttaattt	acttaacgat
	120181	gattcaaaata	tagttaaaaca	aggtttaaatg	tgaatggagc	aatacgcctat	ctataataaa
	120241	gctgtatgat	tcaatgaatg	taatcgaaca	aatctaataa	ttacgaatgg	agcatacaac
	120301	tatgaaaatg	gcagcaattg	cgaaagcaag	tttagcatta	ggtatttttag	caacaggaac
	120361	aataacgtca	ttgcatcaaa	ctgtaaatgc	gagtgaaacat	gaagcaaaaat	atgaaaatgt
40	120421	gacaaaagat	atctttgact	taagagatta	ctatagtggc	gcaagtaagg	aacttaaaaa
	120481	tgttactggg	tatcgttata	gcaaaggtgg	caagcattac	cttatctttg	ataaaaaatag
	120541	aaaattcaca	agagtacaga	tatttggtaa	agatattgaa	agattttaag	cacgcaaaaa
	120601	tccggggatta	gacataattg	ttgttaaaga	agcagaaaaat	cgcaacggga	cagtgttttc
	120661	atatgggtgg	gtcactaaga	aaaactcaaga	cgcttattat	gattatataa	acgcaccaag
45	120721	atttcaaatc	aagagagatg	aaggtgacgg	tattgctacg	tacggtagag	tacactacat
	120781	ttataaagaa	gagatttcac	ttaaagaact	cgacttttaa	ttgagacagt	atttaattca
	120841	aaattttgat	ctgtataaaa	agtttcctaa	agatagtaag	ataaaaagta	taatgaaaga
	120901	tggcgggctat	tatacgtttg	aacttaataa	aaaattacaa	acaaatcgca	tgagtgatgt
	120961	cattgacggg	agaaaatattg	aaaaaataga	agccaacatt	agataattca	atgaaatatg
50	121021	gataaatagta	aaatatggat	agtatagagg	agttaggcaa	cataagttgc	ttagcttctt
	121081	ttttgtgttg	gcgagatgaa	aatgaagcgt	atcgatgaat	aataaaaaaca	ccaataaaaac
	121141	ttgtggaaat	agttgatact	tatagatcgc	tgatgtcgct	ttagtgacat	gaaacaatgt
	121201	ggaaaacata	attaaattga	gggaaagtgt	gaatagttaa	aaaatttagca	ttgtgttata
	121261	aaaaataatt	aatactgtta	ggattttcatt	aactaactta	acgttgggtt	aaaaatagtt
55	121321	aaaaagagggt	taattcattag	cgcagtatct	cgcttatata	atgatatagat	attgttcgta
	121381	ttacgtaatt	gaattaatca	tataaaaaata	tattaagaca	aaattttataa	atagattggg
	121441	agaatagtac	tatgaaatta	aaaacgtag	ctaaagcaac	attggcatta	ggcttattaa
	121501	ctactgggtg	gattacatca	gaaggccaag	cagtccacgc	aaaagaaaaag	caagagagag
	121561	tacaacattt	atatgatatt	aaagacttat	atcgatacta	ctcatcagaa	agttttgaat
60	121621	tcagtaatat	tagtggtaag	gttgaaaaact	ataacggttc	taacgttgta	cgctttaacc
	121681	aagaaaaaca	aaatcaccaa	ttattcttat	taggaaaaaga	taaagataaa	tataaaaaag
	121741	gccttgaagg	ccagaatgtc	tttgtggtaa	aagaattaat	tgatocaaac	ggtagactat
	121801	ctactgttg	tggtgtgact	aagaaaaata	acaaatcttc	tgaaactaat	acacatttat
	121861	ttgttaataa	agtgtatggc	ggaaatttag	atgcatcaat	tgactcattt	ttaatttaata
65	121921	aagaagaagt	ttcactgaaa	gaacttgatt	tcaaaattag	aaagcaatta	gttgaaaaat
	121981	atgggtttata	taaaaggtacg	actaaatacgc	gtaagatcac	tatcaatttg	aaagacgaga

	122041	aaaaggaagt	aattgattta	ggtgataaac	tgcaattcga	gcgcatgggt	gatgtgttga
	122101	atagtaagga	tattcaaaat	atagcagtga	ctattaatca	aatttaaat	aaacaatcaa
	122161	tgacttttaa	gtaataaatt	tgaagcagct	taacgatgaa	atgttgaaata	aatacgtaca
5	122221	tcctccaaaa	aggggcgat	ctaaatcaac	agtgtcgtta	ggctgttttt	atgttttata
	122281	acgcagggtta	tgagcgtact	aaaaattcac	attacttctg	aaagtgtatg	ccattgaata
	122341	ttaattagtt	cttcattaac	catgatttaa	ttttaattaa	acgagtgtta	atgtcagttc
	122401	gtctcaatgc	cctttataat	aaatgtgtat	tattcaaat	acgtaataaa	agcaatccaa
	122461	tatattaaga	ttggagcata	tgaatatgaa	atttacagcg	atagctaaa	cgatatttgt
10	122521	attaggaata	ttacaacaaa	gtgtaattgat	aacagaaaa	caatcgggtta	atgcaaaaagg
	122581	aaagtatgaa	aaaatgaacc	gtttatatga	tacaaacaag	ttacatcaat	actattcagg
	122641	acctagttat	gagttaacaa	atgttagtg	ccaaagtcaa	ggttattatg	actctaacgt
	122701	tttgcttttt	aaccaacaaa	atcaaaagtt	ccaagtattt	ttattgggaa	aagatgaaaa
	122761	taaatcaaaa	gaaaaaacac	atggttttaga	tgtctttgcg	gtaccagaat	tagtagattt
	122821	agatggaaga	atatttagtg	ttagtgtgtg	aacgaagaaa	aacgtaaaat	caatatttga
15	122881	gtctctaaga	acgccgaact	tactagttaa	aaaaatagac	gataaagacg	gtttttctat
	122941	tgatgaattt	ttctttattc	aaaaggaaga	agtgtcattg	aaggaaactg	attttaaaat
	123001	aagaaaactg	ttgattaaaa	aatacaaaact	gtatgaagg	tcagctgata	aaggtagaat
	123061	tgttattaat	atgaaagatg	aaaaataagta	tgaaattgat	ttaagtata	aatttagattt
	123121	cgagcgtatg	gcagatgtca	ttaatagtga	acaaattaag	aacatcgaag	tgaatttgaa
20	123181	ataatcaatg	atataataga	taaaagctta	agaagcgggt	taataatccc	atgtttaatg
	123241	attttgatac	gtgttttaat	aataaaaaca	tatcgaacat	tgactacgtt	attaagctgc
	123301	ttttttgtac	actttgtatc	gaataaactta	agatctaaaa	ctaactcgaa	agaacaatga
	123361	ttcccccaaa	aaaatttatg	ttgctattaa	aaatcagtta	atacgaatgt	taaaatacgt
25	123421	ttgattttca	ttataaatga	ttcaagttta	tttaaatgag	cgttaatgtc	agctgttttt
	123481	gatgcacott	ataataaaga	cagatagttc	aaattacgta	atcatacaaa	tccaatatat
	123541	caagattgga	gcaaataaat	atgaaattga	ctgcattagc	aaaagcaaca	ttagcattag
	123601	gaatattaac	tacaggtgtg	tttacagcag	aaagtaaagc	tgttcacgcg	aaagtagaac
	123661	ttgatgagac	acaacgcaaa	tattatatca	atatgctaca	tcaatactat	tctgaagaaa
30	123721	gttttgaatc	aacaaaacatt	agcgttaaaa	gtgaagatta	ctatggttct	aacgttttaa
	123781	actttaacca	acgaaataaa	actttcaaa	tggttctgct	tggtgacgat	aaaaataaat
	123841	ataaagaaaa	aacacatggc	cttgatgtct	ttgcagtacc	tgaattaata	gatataaaag
	123901	gtggcatata	tagcgttggc	ggtataacaa	agaaaaatgt	gagatcagtg	tttggatttg
	123961	taagtaatcc	aagtctacaa	gttaaaaaag	ttgatgctaa	acatggcttt	tcgataaatg
35	124021	agttgttctt	tattcaaaag	gaagaagtat	cgttgaagga	actggatttg	aaaataagaa
	124081	aaatgtttagt	cgaaaaatat	agattgtata	aaggcgcgtc	agataaaagt	agaatcgtta
	124141	ttaatatgaa	agacgaaaa	aaatatgtaa	ttgatttaag	tgaaaaatta	agttttgatc
	124201	gtatgtttga	tgtaattgat	agtaagcaaa	ttaaaaatat	tgaagtgaat	ttgaattagt
	124261	ttgagttaat	agcataatag	cttaagaagc	gacttaacga	caaaatgtga	attgacatct
40	124321	gtgtccttat	ataaggaact	gtgttaataa	cattactgtt	gttaagttgt	ttttgtaatt
	124381	caaagagcag	aacagagtaa	catcatcagt	tgtagtaaac	gataatccag	taaaacaact
	124441	aaatgaaata	atgaaagtca	tttaacctga	acattaaaa	atatttgttt	ttcatthaaga
	124501	ataattcaag	tatattttaa	tcgaggttaa	ttatcgtatg	aaacgatgca	cgttataata
	124561	aaaatgtatg	attcaaatga	cgtaattgaa	acaatccaat	atattaagat	tgagagcaaat
45	124621	aaatatgaaa	tttacagcat	tagcaaaagc	aacattagct	ttaggaattt	taacaacagg
	124681	aactttaaca	acagaagttc	attcaggtca	tgcaaaaaca	aatcaaaagt	cagtaataaa
	124741	acatgacaag	gaggcactat	accgatcta	cactggaaa	actatggaaa	tgaaaaatat
	124801	tagtgctttg	aaacatggta	aaaataactt	gcgttttaag	tttagaggta	ttaagattca
	124861	agttttactg	cctggaaatg	ataaaaagtaa	atttcaacag	cgtagttatg	aggggttaga
50	124921	tgtgtttttt	gttcaagaaa	aaagagataa	gcacgatata	ttttatactg	ttggtgtgtg
	124981	aaacagaaat	aataaaaacat	ctggagtgtg	cagtgcacca	atattaaata	tttcaaaaaga
	125041	aaagggtgaa	gatgctttttg	tgaagggtta	cccttattac	attaaaaaag	aaaaataaac
	125101	actaaaagag	ctggattata	agttgagaaa	gcattctaatt	gaaaaatacg	gactttataa
	125161	aacaatctca	aaagatggta	gggtcaaaat	tagcttgaaa	gatggcagtt	tttataacct
55	125221	tgattttaaga	tctaaattaa	aatttaaaata	tatgggggaa	gtcatagaaa	gcaaacaaat
	125281	taaagatatt	gaagttaact	taaagtaaat	aattacgaat	aataaaaaag	aattgaagcg
	125341	gcttaacggt	gaaatgtaaa	ttggtgcgca	tagcttatat	aaaaaggatg	catcaatcga
	125401	tatcgtcgtt	aagccgtttt	tggttctgtg	tcattgaatcc	tatcccaatc	tccatgaata
	125461	taaaaattcc	accaccaaca	tcaaaattct	ccacatcgca	acataacca	atgttataat
60	125521	aaatctatta	cacaaagaga	taaaattact	attcaaaagg	ggaggaatca	catgtctatt
	125581	actgaaaaac	aacgtcagca	acaagctgaa	ttacataaaa	aattatgggtc	gattgcgaat
	125641	gatttaagag	ggaatttgga	tgcgagtga	ttccgtaatt	acatttttag	cttgattttc
	125701	tatcgcttct	tatctgaaaa	agcggaacaa	gaatatgcag	atgccttgct	agtggaagat
	125761	attacgtatc	aagaagcatg	ggcagatgaa	gaatatcgtg	aagacttaaa	agctgaatta
65	125821	attgatcaag	tcggttactt	cattgaacca	caagatttat	tcagtgcgat	gattcgtgaa
	125881	attgaaacgc	aagatttcga	tatcgaacat	ctcgcacagg	cgattcgtaa	agttgaaaca

125941 tcaacactag gtgaagaaag tgaaaatgac tttatcggac tgttcagcga tatggactta  
 126001 agttcaacgc gactaggttaa caatgtcaaa gaacgtactg cactaatttc caaagttatg  
 126061 gtttaaccttg acgacttacc attcgttcac agtgatatgg aaattgatat gtttaggtgat  
 126121 gcatacgaat ttcttatcgg gcgctttgcg gcgacagcag gtaaaaaagc aggcgagttc  
 5 126181 tatacaccac aacaagtatc taagatactg gcgaagattg tcacagacgg taaagataaa  
 126241 ttacgtcacg tgtatgaccc aacatgtggt tcaggttcat tattgttacg tgttggttaa  
 126301 gaaacgcaag tgtatcgtaa tttcgggtcaa gaacgtaaca ataccactta caacttagcg  
 126361 cgcataaaca tgttattaca tgatgtgcgc tatgaaaact ttgatatccg taatgatgat  
 126421 acgttggaata atccagcttt cttaggcaat acatttgatg cggttatagc gaaccaccca  
 10 126481 tacagtgcga aatggactgc agattcaaaa tttgaaaatg atgaacgctt cagtggttac  
 126541 ggcaagcttg cgccaaaatc caagcagac tttgccttta ttcaacatag ggtacattac  
 126601 ctagacgatg aaggtaccat ggcagttgta ctccacatg gtgtcttatt ccgtggtgcc  
 126661 gcagaaggcg tcattcgtcg ttattttaatt gaagaaaaga actacttaga agcgtgatt  
 126721 ggcttaccag tgaatatttt ctatgggaca agtattccaa catgtatctt agtattttaa  
 15 126781 aaatgttgcc aacaagacga caacgtatta tttatcgatg catccaatga ttttgaaaaa  
 126841 ggaaaaaatc aaaaccattt aagcgtatgcc caagtcgaac gtattattga cacatacaag  
 126901 cgtaaagaaa caattgataa atacagttac agtgcgacat tacaagagat tgccgataac  
 126961 gattacaacc taaacatacc gaggtatgtc gatacattcg aagaagaagc gccaatgat  
 127021 ttagatcaag tccaacaaga tttgaaaaat atcgacaaag aaatcgcaga aattgaacaa  
 20 127081 gaaatcaatg catacctgaa agaacttggg gtgttgaaag atgagtaata cacaaaaaga  
 127141 aaatgtgcca gagttgagat tcccagggtt tgaaggcgaa tgggaagaga agcagttggg  
 127201 ggatcttaca gatagagtaa ttaggaaaaa taaaaactta gaatcgaaaa agcctttaac  
 127261 aatatccgga cagttaggtt taattgatca aacagaatat tttagtaaat cagtttcgtc  
 127321 gaaaaatcta gaaaattata cactaataaa gaatggagaa ttcgcgataa acaaaagtta  
 25 127381 ttctaattgga taccatttag gggctattaa aagattaact agatatgata gtggtgtatt  
 127441 gtcctctttg tatatttgtt tttctattaa aagtgaatg tctaaagact tcatggaagc  
 127501 atattttgat tcgacacact ggtatagaga agtttctgga attgcagttg aggggcaag  
 127561 aaatcacgga ttattaaatg tttctgtgaa tgattttttt actattctaa ttaaatatcc  
 127621 aagtttagaa gaacagcaaa aaataggcaa gttcttcagc aaactcgacc gacaaattga  
 30 127681 attagaagaa caaaagcttg aattacttca acaacagaaa aaaggctata tgcagaaaat  
 127741 tttctcacag gaactgcgat tcaaagatga gaatggtgaa gattatccag attgggaaaa  
 127801 tagcaaaaata gaaaaatatt taaaagagag aaacgaacgt tctgacaaag ggcaaatgct  
 127861 ttcagtaact ataaatagtg gcatataaaa atttagtgaa ttggatagaa aagataattc  
 127921 aagtaaagat aaaagtaatt ataaagtagt taggaaaaat gatattgcat ataattctat  
 35 127981 gagaatgtgg caaggggcta gtggtaaatc aaattataat gggattgtta gccctgcata  
 128041 tactgtgctt tatccaacac aaaatactag ctcatatttt attggatata agtttaaaac  
 128101 acatagaatg attcataaat ttaaaattaa ttcacaagga ttaacatcag atacatggaa  
 128161 cttaaaatat aaacaattaa aaaatataaa tatagatata cctgtattgg aggaacaaga  
 128221 aaagataggt gatttcttta aaaaaatgga tatattgata agtaaacaga aaatgaaaaat  
 40 128281 tgaaatatta gaaaaagaga aacaatcctt tttacaaaaa atgttcttat aactttgata  
 128341 aatacataga ttgcataaga ataaaatttg tataatttaa cataaaagtt gtaaaagtaa  
 128401 agtgaattaa aaacgaacat taaatttagg cactgtgaaa gcgcagtgct ttttttgtgt  
 128461 cgaaattgtg tacagaataa gtatgttaaa aaagattaag ttgagataaa gtgttattcg  
 128521 taaataaaaag agagtagatc gataggaaat gaatgatatt agttaactat ttattaaatt  
 45 128581 acttaataat gattaatttt tagttaaagt aagtttaatg tgaagcacga ccattgctca  
 128641 ttataatgaa tgaggattgt tcgtattgcg taatagaata aatcaaatag actaaaaatt  
 128701 gggagcatag aattatgaaa ttaaaaaata ttgctaaagc aagtttagca ctagggtatt  
 128761 taacaacagg tatgattaca actactgctc agccagttaa agcaagtaca tttagggtta  
 128821 gatcacaagc tactcaagac ttgagtgaat attataaagg gagaggattt gaattaacaa  
 50 128881 atgtaacagg atataaatac ggaaataaag ttacattttat agataattct caacaaattg  
 128941 atgttacatt gacaggaaat gaaaaattaa ctgttaaaga tgatgacgaa gtttctaatt  
 129001 ttgacgtgtt ttagtaaga gaaggtagtg acaaatcagc tatcacaaca tcgattgggtg  
 129061 gaattacaaa gacaaatggg actcaacata aagatactgt tcaaaacggt aatttgtcag  
 129121 tttctaagag tacagggtcaa cacactactt ctgtgacttc agaatactat agcatctata  
 55 129181 aagaagaaat ctcatataaa gaacttgatt ttaaaattaag aaagcattta attgataaac  
 129241 atgatcttta taagacagaa cctaagagaca gtaaaattag aattactatg aaaaacgggtg  
 129301 gctattatac atttgaatta aataaaaaat tacagcctca tcgcatgggt gatacgattg  
 129361 atagtagaaa tatagagaaa attgaagtga atttataata atattcgagg gagtatatca

60 SEQ ID No:17- SSL1

65 Gene /gene="(set6) Mu50ss11"  
CDS 116916..117596  
 /gene="(set6)Mu50ss11"

5                    /note="Pathogenicity island SaPin2  
SAV0422"  
                  /codon\_start=1  
                  /transl\_table=11  
                  /product="(exotoxin 6) SSL1"  
                  /protein\_id="BAB56584.1"  
                  /db\_xref="GI:14246190"

10           /translation="MKFKAIKASIALGMLATGVITSNVQSVQAKTEVKQQSESELKH  
YYNKPVLERKNVTGYKYTEKGKDYIDVIVDNQYSQISLVGSDKDKFKDGDNSNIDVFI  
LREGDSRQATNYSIGGVTKTNSQPFIDYIHTPILEIKKGKEEPQSSLYQIYKEDISLK  
ELDYRLRERAIKQHGLYSNGLKQGQITITMKDGKSHITIDLSQKLEKERMGDSIDGRQI  
QKILVEMK"

15    SEQ ID No:18 - SSL2

gene                    117882..118577  
                                          /gene="(set7) Mu50ss12"  
                  CDS                    117882..118577  
20                    /gene="(set7) Mu50ss12"  
                                          /note="Pathogenicity island SaPin2  
SAV0423"  
                                          /codon\_start=1  
                                          /transl\_table=11  
25                    /product="(exotoxin 7) SSL2"  
                                          /protein\_id="BAB56585.1"  
                                          /db\_xref="GI:14246191"

30           /translation="MKMKNIKISLLLGLATGVNTTTEKPVHAEKKPIVISENSKKL  
KAYYTQPSIEYKNVTGYISFIQPSIKFMNIIDGNSVNNLALIGKDKQHYHTGVHRNLN  
IFYVNEDKRFEGAKYSISGITSANDKAVDLIAEARVIKADHIGEYDYDFPFKIVKEA  
MSLKEIDFKLRKYLIDNYGLYGEMSTGKITVKKKYKGKTYFELDKKLQEDRMSDVINV  
TDIDRIEIKVRKA"

35    SEQ ID No:19 - SSL3

gene                    118868..119938  
                                          /gene="(set8) Mu50ss13"  
40                    CDS                    118868..119938  
                                          /gene="(set8) Mu50ss13"  
                                          /note="Pathogenicity island SaPin2  
SAV0424"  
                                          /codon\_start=1  
45                    /transl\_table=11  
                                          /product="(exotoxin 8) SSL3"  
                                          /protein\_id="BAB56586.1"  
                                          /db\_xref="GI:14246192"

50           /translation="MNMKTIAKTSLALGLLTTGAITVTTQSVKAEKIQSTKVDKVP  
TLKAERLAMINITAGANSATTQAANTROERTPKLEKAPNTNEEKTSAKIEKISQPKQEE  
QKSLNISATPAPKQEQSQTTESTTPKTKVTPPSTNTPQPMQSTKSDTPQSPTIKQA  
QTDMPKYEDLRAYYTKPSFEFEKQFGFLLKPWTTVRFMNVIPNRFIYKIALVGKDEK  
KYKDGPDYDIDVFIIVLEDNKYQLKKYSVGGITKTNSKKVDHKAELSVTKKDNQGMISR  
55           DVSEYMITKEEISLKEIDFKLRKQLIEKHNLGNGMSGTIVIKMKNNGGKYTFELHKKL  
QEHMADVIEGTNIDKIEVNIK"

SEQ ID No:20 - SSL5

60           gene                    120302..121006  
                                          /gene="(set10) Mu50ss15"  
                  CDS                    120302..121006  
                                          /gene="(set10) Mu50ss15"  
                                          /note="Pathogenicity island SaPin2]  
65                    SAV0425"

```

5          /codon_start=1
          /transl_table=11
          /product="(exotoxin 10) SSL5"
          /protein_id="BAB56587.1"
          /db_xref="GI:14246193"

/translation="MKMAAIAKASLALGILATGTITSLHQTVNASEHEAKYENVTKDI
FDLRDYYSGASKELKNVTGYRYSKGGKHYLIFDKNRKFTRVQIFGKDIERFKARKNPG
LDIFVVKEAENRNGTVFSYGGVTKKNQDAYDYINAPRFQIKRDEGDGIATYGRVHYI
10 YKEEISLKELDFKLRQYLIQNFDLYKKFKPKDSKIKVIMKDGYYTFELNKKLQTNRMS
DVIDGRNIEKIEANIR"

SEQ ID No:21 - SSL7

15     gene      121452..122147
          /gene="(set11) Mu50ssl7"
          CDS      121452..122147
          /gene="(set11) Mu50ssl7"
          /note="Pathogenicity island SaPIIn2
20 SAV0426"
          /codon_start=1
          /transl_table=11
          /product="(exotoxin 11) SSL7"
          /protein_id="BAB56588.1"
25          /db_xref="GI:14246194"

/translation="MKLKTAKATLALGLLTTGVITSEGGQAVHAKKQERVQHLYDIK
DLRYRYSSESEFSNISGKVENYNGSNVVRFNQEKQNHQLFLLGKDKDKYKKGLEQON
VFVVKELIDPNGRLSTVGGVTKKNNKSSETNTHLFVNKVYGGNLDASIDSFLINKEEV
30 SLKELDFKIRKQLVEKYGLYKGTTKYKGITINLKDEKKEVIDLGDKLQFERMGDVLS
KDIQNIATVINQI"

SEQ ID No:22 - SSL8

35     gene      122486..123184
          /gene="(set12) Mu50ssl8"
          CDS      122486..123184
          /gene="(set12) Mu50ssl8"
          /note="Pathogenicity island SaPIIn2
40 SAV0427"
          /codon_start=1
          /transl_table=11
          /product="(exotoxin 12) SSL8"
          /protein_id="BAB56589.1"
45          /db_xref="GI:14246195"

/translation="MKFTAIAKAIFVLGILTTSVMITENQSVNAKGKYEKMNRLYDTN
KLHQYYSGPSYELTNVSGSQGYDSNVLLFNQQNQKFQVFLGKDKENKYKEKTHGLD
50 VFAVPELVLDLGRIFSVSGVTCKNVKSIFESLRTPNLLVKKIDDKDGFSEFFFIQK
EEVSLKELDFKIRKLLIKKYKLYEGSADKGRIVINMKDENKYEIDLSDKLDFERMADV
INSEQIKNIEVNLK"

SEQ ID No:23 - SSL9

55     gene      123561..124259
          /gene="(set13) Mu50ssl9"
          CDS      123561..124259
          /gene="(set13) Mu50ssl9"
          /note="Pathogenicity island SaPIIn2
60 SAV0428"
          /codon_start=1
          /transl_table=11
          /product="(exotoxin 13) SSL9"
65          /protein_id="BAB56590.1"

```

/db\_xref="GI:14246196"

/translation="MKLTALAKATLALGILTTGVFTAESKAVHAKVELDETQRKYYIN  
MLHQYYSEESFESTNISVKSEDIYGSNVLNFNQNRKTFKVFLLGDDKNKYKEKTHGLD  
VFAVPELIDIKGGIYSVGGITKKNVRSVFGFVSNPSLQVKKVDAKHGFSINELFFIQK  
EEVSLKELDFKIRKMLVEKYRLYKGDSDKGRIVINMKDEKKYVIDLSEKLSFDRMFDV  
MDSKQIKNIEVNLN"

SEQ ID No:24 - SSL10

gene 124625..125308  
/gene="(set14) Mu50ss110"  
CDS 124625..125308  
/gene="(set14) Mu50ss110"  
/note="Pathogenicity island SaPin2  
SAV0429"  
/codon\_start=1  
/transl\_table=11  
/product="(exotoxin 14)SSL10"  
/protein\_id="BAB56591.1"  
/db\_xref="GI:14246197"

/translation="MKFTALAKATLALGILTTGTLTTEVHSGHAKQNQKSVNKHDKA  
LYRYTGTGKTMEMKNISALKHGKNNLRFKFRGIKIQVLLPGNDKSKFQQRSYEGLDVFF  
VQEKRDKHDIFYTVGGVIQNNKTSQVVSAPILNISKEKGEDAFVKGYPYIKKEKITL  
KELDYKLRKHLIEKYGLYKTISKDGRVKISLKDGSFYNLDLRSKLKFKYMGEVIESKQ  
IKDIEVNLK"

SEQ ID No:25 - SSL11

gene 128715..129398  
/gene="(set15) Mu50ss111"  
CDS 128715..129398  
/gene="(set15) Mu50ss111"  
/note="Pathogenicity island SaPin2  
SAV0433"  
/codon\_start=1  
/transl\_table=11  
/product="(exotoxin 15)SSL11"  
/protein\_id="BAB56595.1"  
/db\_xref="GI:14246201"

/translation="MKLKNIKASLALGILTTGMITTTAQPVKASTLEVRSQATQDLS  
EYYKGRGFELTNVTGYKGNKVTIDNSQQIDVTLTGNEKLTVKDDDEVSNVDVFFVR  
EGSDKSAITTSIGGITKTNGTQHKDTVQNVNLSVSKSTGQHTTSVTSEYYSIYKEEIS  
LKELDFKLRLKHLIDKHDLYKTEPKDSKIRITMKNGGYTTFELNKKLQPHRMGDTIDSR  
NIEKIEVNL"

**S. aureus strain MW2 taken from GenBank**

SEQ ID No:26 - SSL1

139029..139709

CDS 139029..139709  
/gene="(set16)MW2ss11"  
/gene="(set16) MW2ss11"  
/note="ORFID:MW0382  
exotoxin homolog [Genomic island nu Sa alpha2]"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB94247.1"  
/db\_xref="GI:21203547"

/translation="MKFKAIKASLALGMLATGVITSNVQSVQAKTEVKQQSEADLKL

YYNGPSFEYKKVGTGYGFIEGKDRFIDFIYNGQYNKISLVGSDKDKYNEEVNPDIDVFW  
VREGNGRQADNHSIGGITKTNRGVYDYIHTPILEIKKGKEEPQSSLYQIYKEDISLK  
ELDFKLRKQLISQSGLYSNGLKQGQITITMNDGTTHTIDLSQKLEKERMGESIDGRQI  
QKILVEMK"

5

SEQ ID No:27 - SSL2

gene 139995..140690  
/gene="(set17) MW2ss12"  
10 CDS 139995..140690  
/gene="(set17) MW2ss12"  
/note="ORFID:MW0383  
exotoxin homolog [Genomic island nu Sa alpha2]"  
/codon\_start=1  
15 /transl\_table=11  
/protein\_id="BAB94248.1"  
/db\_xref="GI:21203548"  
  
/translation="MKMKSIVKISLLLGILATGVNTTTEKPVHAEKKPIVISENSKKL  
20 KAYYTQPSIEYKQVNTGYISFIQPSIKFMNIIDGNSVNNIALIGKDKQHYHTGVHRNLN  
IFYVNEDKRFEKAKYSIGGITSANDKAVDLIAEARVIKADHIGEYDYDFFPFKIDKEA  
MSLKEIDFKLRKYLDNYGLYGEKSTGKITVKKKYGYKTYFELDKKLQEDRMSDVINV  
TDIDRIEIKVRKA"

25

SEQ ID No:28 - SSL3

gene 140981..142051  
/gene="(set18) MW2ss13"  
30 CDS 140981..142051  
/gene="(set18) MW2ss13 "  
/note="ORFID:MW0384  
exotoxin homolog [Genomic island nu Sa alpha2]"  
/codon\_start=1  
/transl\_table=11  
35 /protein\_id="BAB94249.1"  
/db\_xref="GI:21203549"  
  
/translation="MKMRTIAKTSLALGLLTTGATVTTQSVKAEKIQSTKVDKVPVL  
40 KAERLAMINITAGANSATTQAANTROERTPKLEKAPNTNEEKTSASKIEKISQPKQEE  
QKTLNISATPAPKQEQSQTTTESTTQQTQMTTPPSTNTPQPMQSTKSDTPQSPTIKQA  
QTDMPKPYEDLRAYYTKPSFEFEKQFGFLKPPWTTVRFMNVIPNRFIYKIALVGKDEK  
KYKDGPDYDNIDVFIVLEDNKYQLKKYSVGGITKTNSKKVNHKVELSITKKDNQGMISR  
DVSEYMITKEEISLKEIDFKLRKQLIEKHNLGYNGMSGTIVIKMKNNGGKYTFELHKKL  
45 QEHRMADVIDGTNIDNIEVNIK"

45

SEQ ID No:29 - SSL4

gene 142416..143363  
/gene="(set19) MW2ss14 "  
50 CDS 142416..143363  
/gene="(set19) MW2ss14"  
/note="ORFID:MW0385  
exotoxin homolog [Genomic island nu Sa alpha2]"  
/codon\_start=1  
55 /transl\_table=11  
/protein\_id="BAB94250.1"  
/db\_xref="GI:21203550"  
  
/translation="MKITTIKTSALGLLTTGVITTTTQAANATTPPSTKVETPQQV  
60 ANATTPSSTKVEAPQQAANATTPSSTKVEAPQSKPNATTPSSTKVEAPQQAANATTP  
SSNVDTSPPPQSPPTTKQVPTEINPKFKDLRAYYTKPSLEFKNEIGIILKKWTTIRFMNV  
VPDYFIYKIALVGKDDKKYGEVHRNVDFVVLLENNYNLEKYSVGGITKSNSKKVDH  
KAGVRITKEDNKGITSHDVSEFKITKEQISLKEIDFKLRKQLIEKNNLYGNVSGKIV  
IKMKNNGGKYTFELHKKLQENRMADVIDGTNIDNIEVNIK"

65

SEQ ID No:30 - SSL5

5      gene                      143727..144431  
          /gene="set20 MW2ss15"  
          CDS                      143727..144431  
          /gene="(set20) MW2ss5 "  
          /note="ORFID:MW0386  
          exotoxin homolog [Genomic island nu Sa alpha2]"  
          /codon\_start=1  
          /transl\_table=11  
          /protein\_id="BAB94251.1"  
          /db\_xref="GI:21203551"

10

15      /translation="MKMAAIKASLALGILATGTTTSLHQTVDNASEHEAKYENVTKDI  
 FDLRDYYSGASKELKNVTGYRYSKGGKHYLIFDKHQKFTRIQIFGKDIERFKARKNPG  
 LDIFVVKAEANRNGTVFSYGGVTCKNQDAYDYINAPRFQIKRDEGDGIATYGRVHYI  
 YKEEISLKELDFKLRQYLIQNFDLYKKFKPKDSKIKVIMKDDGGYYTFELNKKLQTNRMS  
 DVIDGRNIEKIEANIR"

20      SEQ ID No:31 - SSL6

gene                      144877..145575  
          /gene="(set21) MW2ss16"  
          CDS                      144877..145575  
          /gene="(set21) MW2ss16"  
          /note="ORFID:MW0387  
          exotoxin homolog [Genomic island nu Sa alpha2]"  
          /codon\_start=1  
          /transl\_table=11  
          /protein\_id="BAB94252.1"  
          /db\_xref="GI:21203552"

25

30      /translation="MKLKALAKATLVLGLLATGVITTESQTVKAAESTQGQHNYKSLK  
 YYYSKPSIELINVDGLYRQHLTDKGAYVWKNLKDYYIGLLGEDSKKFKSDVYGDLDFAF  
 LVIEEEPVKGRQYSIGGISKTNSKEFKEREVDVKVTRKADRDTTSTKDSKFKITKEEI  
 SLKELDFKLRQKLMKEENLYDAINHRKGKIVVKMEDDKFYTFELTKKLQPHRMGDTID  
 GTKIKEINVELEYK"

35

40      SEQ ID No:32 - SSL7

gene                      145997..146692  
          /gene="(set22)MW2ss17"  
          CDS                      145997..146692  
          /gene="(set22) MW2ss17"  
          /note="ORFID:MW0388  
          exotoxin homolog [Genomic island nu Sa alpha2]"  
          /codon\_start=1  
          /transl\_table=11  
          /protein\_id="BAB94253.1"  
          /db\_xref="GI:21203553"

45

50      /translation="MKLKTAKATLALGLLTTGVITSEGQAVQAKEKQERVQHLYDIK  
 DLHRYYSSESFEFSNISGKVENYNGSNVVRFNQENQNHQLFLSGKDKDKYKEGLEQON  
 VFVVKELIDPNGRLSTVGGVTCKNNQSSETNTPLFITKKVYGGNLDASIESFLINKEEV  
 SLKELDFKIRQHLVKNYGLYKGTTKYKGKITFNLKDGEKQEI DLGDKLQFEHMGDVLNS  
 KDIQNIAVTINQI"

55

60      SEQ ID No:33 - SSL8

gene                      147031..147729  
          /gene="(set23) MW2ss18"  
          CDS                      147031..147729  
          /gene="(set23) MW2ss11 "  
          /note="ORFID:MW0389  
          exotoxin homolog [Genomic island nu Sa alpha2]"

65

```

                    /codon_start=1
                    /transl_table=11
                    /protein_id="BAB94254.1"
                    /db_xref="GI:21203554"
5
    /translation="MKFTAIAKAIFVLGILTTSMITENQSVNAKGKYEKMNRLYDTN
KLHQYYSGPSYELTNVSGSQGYDSNVLLFNQQNQKFQVFLGKDENVYKEKTHGLD
VFAVPELVLDLGRIFSVSGVTKKNVKSIFESL RTPNLLVKKIDDKGFSYDEFFFIQK
10 EEVSLKELDFKIRKLLIKKYKLYEGAADKGRIVINMKDENKYEIDLSDKLGFERMADV
INSEQIKNIEVNLK"

    SEQ ID No:34 - SSL9

    gene            148108..148806
15    /gene="(set24) MW2ssl9"
    CDS            148108..148806
    /gene="(set24) MW2ssl9"
    /note="ORFID:MW0390
20    exotoxin homolog [Genomic island nu Sa alpha2]"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB94255.1"
    /db_xref="GI:21203555"

25    /translation="MKFTALAKATLALGILTTGVFTTESKAVHAKVELDETQKYYIN
MLHQYYSEESFEPTNISVKSEDIYGSNVLFNFKQRNKAQVFLGDDKNKYKEKTHGLD
VFAVPELIDIKGGIYSGGITKKNVRSVFGFVSNPSLQVKIDPKHGFSINELFFFIQK
EEVSLKELDFKIRKMLVEKYRLYKGASDKGRIVINMKDEKKYVIDLSEKLSFDRMFDV
30 MDSKQIKNIEVNLN"

    SEQ ID No:35 - SSL10

    gene            149165..149848
35    /gene="(set25) MW2ssl10"
    CDS            149165..149848
    /gene="(set25) MW2ssl10"
    /note="ORFID:MW0391
40    exotoxin homolog [Genomic island nu Sa alpha2]"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB94256.1"
    /db_xref="GI:21203556"

45    /translation="MKLTAIAKAALALGILTTGTLTTEVHSGHAKQNQKSVNKHDKA
LYRYYTGKTMEMKNISALKHKGNNLRFKFRGIKIQVLLPGNDKSKFQQRSYEGLDVFF
VQEKRDKHDIFYTVGGVIQNKTSGVVSAPILNISKEKGEDAFVKGYPYIKKEKITL
KELDYKLRKHLIEKYGLYKTISKDGRVKISLKDGSFYNLDLRSKLKFKYMGIEVIESKQ
IKDIEVNLK"

50    SEQ ID No:36 - SSL11

    gene            153324..154016
55    /gene="(set26) MW2ssl11"
    CDS            153324..154016
    /gene="(set26) MW2ssl11"
    /note="ORFID:MW0394
60    exotoxin homolog [Genomic island nu Sa alpha2]"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB94259.1"
    /db_xref="GI:21203559"

65    /translation="MKLKNIKASLALGILTTGMITTTAQPVKAIEQSRLSVTSKDTQ
ELKKYSGTGYNFQNVSGYREGNKMNIIDGPQLNVVTLTGTDKERFKDDEDEYEGLDVF

```

VVREGSGKHADNISIGGITKTNKNQYKDPVQNVNLLTSKSNQONTASVTSEYYSINKE  
EISLKELDFKLKQLIDKHDLYKTEPKDSKIKVSMKNGGYTTFELNKKLQPHRMGDTI  
DSRNIKKIEVNL"

5

*S. aureus* strain NCTC8325 taken from uncompleted genome  
project at Oklahoma University via <http://pedant.gsf.de>

SEQ ID No:37

10

15

20

25

30

35

40

45

50

55

60

65

70

```
45955 CTGATT GTCTACATTT ATATTAGAGA TAAAGCGTT TTGATTTTCA GATGCTTCTG CTTGTCCGTT TGTC
46025 ATAATA TATAACATTG TAATTGCAAC AGATACTAAA CCGACTTTCA TTTTACGTAA CTTAAAAATTT TCCC
46095 TCATGA TATGCTCCCT CGAATAATTT TATAAATTC CTTCAATTTT TTCTATATTT CTGCCATCAA TAAC
46165 ATCACC CATACGGTGT GTTTGTAACT TTTTATTCAA TTCAAATGTG TAGAACCAC CATCTTTCAT AGTA
46235 ATTCGA ATTTTACTGT CTTTAGGTTT TGCTCTATAA AGGTTATGTT TATCAATTAA ATGCTTTCTT AATT
46305 TAAAT CAAGTCTTT TAATGAAAT TCTTCTTTAT TAATTGTATA TGTAGATGAT GTTGACGTTG ATGT
46375 AACACT ATCGATGTTT TTAGTAATTA TTAATTTTAC ATCTTTTACT TTATCAATAT AATTTGAACC GTTT
46445 GTTTTA GTAATACCAC CAATTGAAGC TGTATTACCA GATCTATCAG AATTTTCTCT TACAACAAAT ATAT
46515 CTACAT TAGAAATATC TTCACCAAAA TTTTGTCTTT CATTCACAGT TAAAGTTACA TCATAAGTT GATA
46585 ATTAGG AGTAAACGTC ACTTTTCCTT CCTCTTTATA TCCTGACTGA TTTGTATAT CAAAGAACGG TCTA
46655 TTATAA TATTCACCTCA AGTCTTGAGT AGCTTGATGAT CTAACCTCTA ATGTAATGCTG TTTTACTGGC TGAG
46725 CAGTAG TTGTAATCAT CCCTGTTGTT AAAATCCCTA GTGCTAAACT TGCTTTAGCA ATATTTTTTA ATTT
46795 CATAAT TCTATGCTCC CAATTTTATG TCTATTTGAT TTATTCTATT ACGCAATACG AACAAATCCTC ATTC
46865 ATTATA ATGAGCAATG GTCGTGCTTC ACATTAAACT TACTTTAACT AAAAATTAAT CATTTATTAAG TAAT
46935 TTAATA AATAGTTAAC TAATATCATT CAATTCCTAT CGATCTACTC TCTTTTATTT ACGAATAACA CTTT
47005 ATCTCA ACTTAATCTT TATTTAACTA CTTATTTCTGT ACACAATTTT GACACAAAAA AGACACTGCG CTTT
47075 CACAGT GCCTAAATTT AATGTTGCTT TTTAATTCAC TTTACTTTTA CAACTTTTAT GTTAAATTAT ACA
47145 ATTTTA TTCTTATGCA ATCTATGTAT TTATCAAAGT TATAAGAACA TTTTTTGTAA AAAGGATTGT TTCT
47215 CTTTTT CTAATATTTT AATTTTCATT TTCTGTTTAC TTATCAATAT ATCCATTTTT TTAAGAAAT CACC
47285 TATCTT TTCTTGTTCC TCCAATACAG GTATATCTAT ATTTATATTT TTTAATTTGT TATATTTTAA GTTC
47355 CATGTA TCTGATGTTA ATCCTTGTTA ATTAATTTTA AATTTATGAA TCATTCTATG TGTTTTAAAC TTAT
47425 ATCCAA TAAATAATGA GCTAGTATTT TGTGTTGGAT AAAGCACAGT ATATGCAGGG CTAACAATCC CATT
47495 ATAAT TGATTTACCA CTAGCCCTT GCCACATTCT CATAGAATTA TATGCAATAT CATTTTTCTT AACT
47565 ACTTTA TAATTAATTT TATCTTTACT TGAATTAATC TTTCTATCCA ATTCACATAA TTTTATAATG CCAC
47635 TATTTA TAGTTACTGA AAGCAATTTG CCTTTCTGAG AACGTTCTGT TCTCTCTTTT AATATATTTT CTAT
47705 TTTGCT ATTTTCCCAA TCTGGATAAT CTTACCAATT CTCATCTTTG AATCGCAGTT CCTGTGAGAA AATT
47775 TTCTGC ATATAGCCTT TTTTCTGTTG TTGAAGTAA TCAAGCTTTT GTTCTTCTAA TTCAATTTGT CGGT
47845 CGAGTT TGCTGAAGAA CTTGCCTATT TTTTCTGTTT CTTCTAACT TGGATATTTA ATTAGAATAG TAA
47915 AAAATC ATTCACAGAA ACATTTAATA ATCCGTGATT TCTTGCAACC TCAACTGCAA TTCCAGAAAC TTCT
47985 CTATAC CAGTGTGTCG AATCAAAATA TGCTTCCATG AAGTCTTTAG ACATTTCACT TTTAATAGAA AAAC
48055 AAATAT ACAAGAGGA CAATACACCA CTATCATATC TAGTTAATCT TTTAATAGCC CCTAATGGGT ATCC
48125 ATTAGA ATAACTTTTG TTATACGCGA ATTCCTCCATT CTTTATTAGT GTATAATTTT CTAGATTTT CGAC
48195 GAAACT GATTACTTAA AATATTCTGT TTGATCAATT AAACCTAACT GTCCGGATAT TGTAAAGGC TTTT
48265 TCGATT CTAAGTTTTT ATTTTTCCTA ATTACTCTAT CTGTAAGATC CCCTAACTGC TTCTCTTCCC ATTC
48335 GCCCTC AACCCCTGGG AACCTCAATT CTGGCACAAT TTTCTTTTGT GTATTACTCA TCTTTCAACA CCCC
48405 AAGTTC TTTTCAGGTAT GCATTGATTT CTTTGCCTAAT TTCTGCGATT TCTTTATCGA TATTTTCAA ATCT
48475 TGTTGG ACTTGATCTA AATCAATCGG TGCTTCTTCT TCGAATGTAT CGACATATCT CGGTATATTT AGGT
48545 TGTAAT CGTTATCGGC AATCTCTTGT AGTGTGCGCT TGTAGCTATA TTTATCAATT GTTTCCTTAC GCTT
48615 ATATGT GTCTATAATA CGTTCGACTT GGGCATCTGT TAAATGATTT TGATTTTTTC CTTTTTCAA ATCA
48685 TTGGAT GCATCGATAA ATAGTACGTT GTCGTCTTGT TGGCGACATT TTTTAAATAC TAAATACAT GTTG
48755 GAATAC TTGTCCCAT AAAAAATATC GCTGGCAAC CAATCACAGC TTCTAAGTAG TTCTTTTCTT CAAT
48825 TAAATA ACGACGAATG ACACCTTCTG CAGCACCTCG GAATAATACA CCATGTGGGA GTACAACGGC CATG
48895 GTACCT TCATCGTCTA GGTAAATGTA CATGTGTTGA ATAAAGGCAA AGTCTGCTTT AGACTTAGGC GCAA
48965 GTTTGC CGTAACCACT GAATCGTTCC TCATTTTCAA ACTTTGAATC TGCAGTCCAT TTCGCACTAT ACGG
49035 TGGGTT CGCAATAACC GCATCAAAATG TATTGCCTAA AAAGGCTGGG TTTTCCAAATG GTTCATCATT ACGG
49105 ATATCG AAGTTCTCAT AACGCACATC ATGTAATAAC ATATTCAATG GTGCTAAGTT GTATGTAGTA TTGT
49175 TACGTT CTTGACCGAA ATAACGATAC ACTTGTGTTT CTTTACCAAC ACGTAACCAAC AGTGAACCTG AACC
49245 ACATGT TGGGTCATAC ACGTGACGTA ATTTATCTTT ACCGTCTGTG ACAATCTTCG CCAGTATCTT AGAT
49315 ACTTGT TGTGGTGTAT AGAATCGGCC TGCTTTTTTA CCCGCTGTG CCGCAAAGCG CCAATTAGG AATT
49385 CATATG CATCACCTAA CATATCAATT TCCATCTCAC TGTGAACGAA TGGTAAGTCG TCAAGATTAA CCAT
49455 GACTTT AGAGATTAAA GCAGTACGTT CTTTGACATT GTTACCTAGT CGCGTTGAAC TCAAATCCAT ATCG
49525 CTGAAC AGACCGATAA AGTCATTTTC ACTTCTCTCA CCTAATGTTG ATGTTTCAAC TTTACGAATT GCCG
49595 TCGCAA GGTGTTTCGAT ATCGAAATCT TGCGTTTCAA TTTTCAAGAT CATCGCACTG AATAAATCTT CTGG
49665 CTCAAT GAAGTAACCG ACTTGGTCAA TTAATTTCTGC TTTTAAAGTCT TCACGGTATT CTTCGCTGCG CCAT
49735 GCTTCT TGATACGTGA TGTCTTCAAC TGACAAGGCA TCTGCATATT CTTGTTCCGC TTTTTCAGAT AAGA
49805 AGCGAT AGAAAATCAA GCCTAAAATG TAATTACGGA ATTCACCTCG ATCCATATTC CCTCTTAAAT CAT
49875 CGCAAT CGACCAT TTTTATGTA ATTCAGCTTG TTGCTGACGT TGTTTTTCAG TAATAGACAT GTGA
49945 TTCTCT CGCCTTTGAA TAAGTAATTT ATCTCTTGT GTAATAGATT TATTATAACA TTTGGTTATG TTGC
50015 GATGTG GAGAATTTTG ATGTTGGTGG TGGAAATTTT ACCTTTATGG AGATTGGGAT AGGATTATG ACAC
50085 ACAAC CAAAACGGCT TAACGACGAT ATCGATTGAT GCATCCTTTT TGTATAAGCT ATGCGCACCA ATTT
```

	50155	ACATTT	CACCGTTAAG	CCGCTTCAAT	TACTTTTTAAT	TATTCGTAAT	GATTACTTTT	AAGTTAACTT	CAAT
	50225	ATCTTT	AATTTGTTTG	CTTCTATGA	CTTCCCCAT	ATATTTAAAT	TTTAATTTAG	ATCTTAAATC	AAGG
	50295	TTATAA	AAACTGCCAT	CTTCAAGCT	AATTTTGACC	CTACCATCTT	TTGAGATTGT	TTTATAAAGT	CCGT
5	50365	ATTTTT	CAATTAGATG	CTTCTCAAC	TTATAATCCA	GTCTTTTTAG	TGTTATTTTT	TCTTTTTTAA	TGTA
	50435	ATAAGG	GTAACCTTTC	ACAAAAGCAT	CTTCACCTT	TTCTTTTGAA	ATATTTAATA	TTGGTGCAC	GACA
	50505	ACTCCA	GATGTTTTAT	TATTCTGTAT	TACACCACCA	ACAGTATAAA	ATATATCGTG	CTTATCTCTT	TTTT
	50575	CTTGAA	CAAAGAAAAC	ATCTAACCCC	TCATAACTAC	GCTGTTGAAA	TTTACTTTTT	TCATTTCCAG	GCAG
	50645	TAAAC	TGAATCTTA	ATACCTCTAA	ACTTAAAACG	TAAGTTGTTT	TTACCATGTT	TCAAAGCACT	AATA
	50715	TTTTTC	ATTTCCATAG	TCTTTCCAGT	GTAGTATCGG	TATAATGCCT	CCTTGTCTATG	TTTATTTACT	GACT
10	50785	TTTGAT	TTTGTGTTTG	ATGACCTGAA	TGAAGTCTTG	TTGTTAAAGT	TCCTGTTGTT	AAAATTCCTA	AAGC
	50855	TAATGT	CGCTTTTGCT	AATGCTGTAA	ATTTTCATAT	TATTTGCTCC	AATCTTAATA	TATTGGATTG	TTTT
	50925	CATTAC	GTAATTTGAA	TCATACATTT	TTATTATAAC	GTGCATCGTT	TCATACGATA	ATTAACCTCG	ATTT
	50995	AAATAT	ACTTGAATTA	TTCTTAATGA	AAAACAAATA	TATTTTAATG	TTCAGGTTAA	ATGACTTTCA	TTAT
	51065	TTCAAT	TAGTTGTTTT	ACCGGATTAT	CGTTTACTAC	AACGTGATGAT	GTTACTCTGT	TCTGCTCTTT	GAAT
15	51135	TACAAA	AACAACCTTA	CAACAGTAAT	GTATTTAACA	CAGTTCCTTA	TATAAGGACA	CAAATGCCAA	TTCA
	51205	CATTTT	TCGTTAAGC	CGCTTCTTAA	GCTATTATGC	AACTAATTCG	AACTAATTCG	AATTCACCTC	AATA
	51275	TTTTTA	ATTTGCTTAC	TATCCATTAC	ATCAAACATA	CGTTCAAAAC	TTAATTTTTT	ACTTAAATCA	ATTT
	51345	CATGCT	TCTTTTCGTC	TTTCATATTG	ATAACAATTC	TACCTTTATC	AGACGTTCCCT	TTATACAATC	TATA
	51415	TTTTTC	GATTAAGAGT	TTTCTTATTT	TAAAGTCCAG	TTCTTCAAT	GATACTTCTT	CCTTTTGAAT	AAAA
20	51485	AACAAC	TCGTTTATCG	AAAAGCCATT	TTTAGCATCA	ACTTTTTTAA	CTTGTAGACT	TGGATTACTT	ACAA
	51555	ATCCAA	ACACTGATCT	CACATTTTTTC	TTTGTATATA	CGCCAACGCT	ATATATGCCA	CCTTTTATAT	CTAT
	51625	TAATTC	AGGTACTGCA	AAGACATCAA	GGCCATGTGT	TTTTTCTTTA	TATTTATTTT	TATCGTCACC	AAGT
	51695	AAAAAT	ACTTTAAAAG	CTTATTTCG	TTGTTTAAAG	TTTAAAACGT	TAGAGCCATA	GTAATCTTCA	CTTT
	51765	TAACAC	TAATGTTTGT	TGGTTCAAAA	CTTCTTCAG	AATAGTATTG	ATGTAGCATA	TTGATATAAT	ATTT
25	51835	GCCTTG	TGTCTCATCA	AGTTCTACTT	TCGCGTGACC	AGTTTGACTT	TCTGCTGTAA	ACACACCTGT	AGTT
	51905	AATATT	CCTAATGCTA	ATGTTGCTTT	AGCTATCGTT	GTTAATTTCA	TATTTTTTTG	CTCCAATCTT	AATG
	51975	TATTGG	ATTGTTATTA	TTACGTAATT	TGAACATCT	GTCTTTATTA	TAAGGTGCAT	CAAAACAGAC	TGAC
	52045	ATTAAC	GCTCATTTAA	ATAAACTTGA	ATCATTATTA	ATGAAAATCA	AACGTATTTT	AACATTCGTA	TTAA
	52115	CTGATT	TTTAATAGCA	ACATAAATAT	TTTTTTGGTG	AATCATTGTT	CTTCCGATT	AGTTTAAAT	CTTA
30	52185	AGCTAT	TGGTTATAAA	GTGTACAAA	AAGCAGTTAG	ATAACGTAGT	CAATGTTTCG	TATGTTTTTA	TTAT
	52255	AAAAAC	ACGTATCAAA	ATCATTA AAC	ATGGGATTAT	TAAACCGCTT	CTTAAGCTTT	CATTCTATAT	ATAT
	52325	CATTGA	TTATTTCAAA	TTCACTTCGA	TGTTTTTAA	TTGTTCACTA	TTAATGACAT	CTGCCATACG	CTCG
	52395	AAATCT	AATTTATCAT	TTAAATCAAT	TTTCATCTT	TTTCATCTT	TCATATTAAT	AACAATCTCA	CCTT
	52465	TATCAG	CTACGCTTTC	ATACAGTTTG	TATTTTAA	TCAACAGTTT	TCTTATTTTA	AAATCAAGTT	CCTT
35	52535	CAATGA	CACTTCTTCC	TTTTGAATTA	AGAAAATTC	ATCAATAGAA	AAACCGTCTT	TATCGTCTAT	TTTT
	52605	TAAACT	AGTAAGTTGC	CGCTTCTTAG	AGACTCAAT	ATTGATTTTA	CGTTTTTCTT	TGTTACACCA	CTAA
	52675	CATCAA	ATATTCTTCC	ATCTAAATCT	ACTAATCTG	TACCCGCAAA	GACATCTAAA	CCATGTGTTT	TTTC
	52745	TTTGTA	TTTATTTTCA	TCTTTTCCCA	ATAAAAACAC	TTGGAACCTT	TGATTTTGT	GGTTAAAAAG	CAAA
	52815	ACGTTA	GAGTCATAAT	AACCTTGACT	TTGGCCACTA	ACATTTGTTA	ACTCATAACT	AGGTCTGAA	TAGT
40	52885	ATTGAT	GTAACCTTGT	TGTATCATAT	AAACGGTTCA	TTTTTTCATA	CTTTCCTTTT	GCATTAACCG	ATTG
	52955	ATTTTC	TGTTATCATT	ACACTTGTG	TTAATATPCC	TAATATAAAT	ATCGCTTTAG	CTATCACTGT	AAAT
	53025	TTTATA	TTTATGTTCT	CCAATCTTAA	TATATTGGAT	TGCTTTTAT	ACGTAATTTG	AATAATACAC	CTTT
	53095	ATTATA	AAGGGCATTG	AGACAGACTG	ACATTAACAC	TCGTTTAAAT	AAACTTAAAT	CATTGTTAAT	GAAG
	53165	AGCTAA	TTAATATTCA	ATGGGCATCA	CTTCAAGAG	CAATGTAAAT	TTTTAGTAGT	CTCACTCATA	AACG
45	53235	CGGTTA	TAAACGTAA	AAAAAAGCCT	AACGACAAGG	TTGTTTTAAA	TACGCTCCTT	AATGTAAGGT	GTAC
	53305	GTATCT	ATTCAACATT	TCATCGTTAA	GCCTGCTTCA	ATTTATTACT	TTAGAGTCAT	TGATTACTTA	CTTT
	53375	AAATTT	GTTTCAAAGT	CACCTCAATC	TTATTAATAT	CCTTACTATT	CAACACATCA	CCCATGCGCT	CGAA
	53445	TTGCAA	TTTATACCTT	AAATCAATTT	CTTGTCTTTC	TCCATCTTTC	AAATGTAGAT	TGCTCTTACC	GTAT
	53515	TTAGTC	GTACCTTTAT	ATAAACCTTA	ATTTTAACT	AAATGTTGTC	TAATTTTGAA	ATCAAGTTCT	TTCA
50	53585	GTGAAA	CTTCTTCTTT	ATTAATTTGAA	AATGAGTCAA	TTGATGCATC	TAAATTTCCG	CCATACACTT	TATT
	53655	AACAAA	TAAATGTGTA	TTAGTTTTCAG	AGATTTTGT	ATTTTCTTCA	GTACACACCAC	CAACAGTAGA	TAA
	53725	CTACCG	TTTGGATCAA	TTAATTTCTT	TACCACAAAG	ACATCTTTGC	CTTCAATGCC	TTCTTTATAT	TTCT
	53795	CTTTAT	CTTTACCTAA	TAAGAATAAT	TGGTGATTTT	GATTTTCTTG	GTTAAAGCGT	ACAACGTTAG	AACC
	53865	GTTATA	ATTTTCAACC	TTACCACATA	TATTAATGAA	TTCAAACTT	TCTGATGAGT	AGTATCGATG	TAAG
55	53935	TCTTTA	ATATCATATA	AATGTTGTAC	TCTCTCTTGC	TTTTCTTTTG	CTTGAACCTG	TTGGCCTTCT	GATG
	54005	TAATCA	CACAGTAGT	TAATAAGCCT	AATGCCAATG	TTGCTTTAGC	TAACGTTTTT	AATTTACACG	TACT
	54075	ATPCTC	CCAATCTATT	TATAAATTTG	TCTTAATATA	TTTTTATATG	ATTAATTCAT	TTACGTAATA	CGAA
	54145	CAATCT	ACTATCATTA	TATAAGTGAG	ATACTGCGCT	ATGAATTAAC	CTCTTTTAA	CTATTTTGA	ACCA
	54215	ACGTTA	AGTTAGTTAA	TGAAATCTTA	ACAGTATTTA	TTATTTTTTA	TAACACAATA	CTAATTTTTT	AAC
60	54285	ATTCAC	ACTTTCCCTC	AATTTAATTA	TGTTTTCCAC	ATTGTTTCAT	GTACGAAAA	GGACACGCG	CGAC
	54355	TATAAG	TATCAACCGT	TTTTCCAAAA	ACTTGATCAT	AAACCCGCTC	CTTTTTTCAT	CATAACAAA	TAAA
	54425	AAAGAA	GCTAAGCAAT	ATGTTATCGC	TTAATTTCCC	TACAATTAAT	AGTCTGCTTG	TCCAAGGATT	ATTT
	54495	ATATTC	TAGCTCAACA	TTAATTTCTT	TGATTTTGGT	ACCATCTATC	GTGTCACCCA	TGCGATGCGG	TTGT
	54565	AGTTTT	TTTGTAAGTT	CGAAAGTATA	AAACTTATCA	TCTTCCATTT	TAACACAAAT	TTTACCTTTT	CTAT
65	54635	TATFAA	CAGCACCATA	TAATTTTCTT	TCTTCCATCA	ATTTTTTTCT	TAATTTAAAG	TCCAACCTCT	TTAA
	54705	CGAGAT	TTCTTCTTTA	TTAATTTTAA	ATTTACTATC	TTTAGACTTT	TCCGATGATT	CATCAATTTT	TCTT
	54775	GTTACT	TTAACATCGA	CTTCTTTACT	AAATCTTTTA	CTATTTGTCT	TACTTAAACC	ACCAATTGAA	TATT
	54845	GTCTTC	CATTAACAGT	TTCTTCTCTG	ATGACTAAAA	ATGCATCTTG	CTTATCATGC	TCACCTTGAG	GGTA
	54915	TTTCTC	AATATCTTTA	CCAAGCAAGC	CACAAAATA	ATCTTTTCGA	TCCTTCCAAA	CATATACTCC	TTTA
70	54985	TCTGTC	ACTTTCTGTC	TATACAAACC	ATCAAGATTT	TTTAACTCTA	TACTTGGCTT	GCTATAGTAG	TATT
	55055	TTAACG	ATTTATAATT	GTGTTGACCT	TGAGTTGATT	CTGCCGCTTT	TACTGTTTGA	CTTTCTGTTG	TTAT
	55125	TACACC	AGTAGCTAAC	AATCCCAATA	CTAATGTTGC	TTTAGCTAAC	GTTTTTAATT	TCATAGTACT	ATTC
	55195	TCCCAA	TCTATTTATA	AATTTTGTCT	TAATATATTT	TTATATGATT	ATTTCAATTA	CGTAATACGA	ACAA

	55265	TCTACT	ATCATTATAT	AAGCGTAATA	CTAAGCTATG	AATTAACCTC	TTTTTAACTA	TTTTTGAACC	AATG
	55335	TTAAGC	TAATTAATGG	AATCCTAACA	GCGTTAATCT	ATTTTTTAAA	CTTAACGCAG	CTTTTTTAAC	TATT
	55405	CACACT	TTCCCTCAAT	TTAACTATGT	TTTTCCACATT	GTTCATGTC	ACGAAAAGGA	CAACGCGCGA	CTAT
5	55475	AAGTAT	CAACTATTTT	CACAAGTTTT	ATTGGTGTGT	TTATTATTCA	TCGATACGCT	TCATTTTCAT	CTCT
	55545	CCAACA	CAAAAAAGAA	GCTAAGCAAC	TTATGTTGCC	TAACCTCTCT	ATACTATCCA	TATTTTACTA	TTAT
	55615	CCATAT	TTGATTGAAT	TATCTAATGT	TGGCTTCAT	TTTTTCAATA	TTTCTACCGT	CAATGACGTC	ACTC
	55685	ATGCGA	TTTGTGTTGA	ATTTTTTAT	AAGTTCAAAC	GTATAATAGC	CGCCATCTTT	CATTATCACT	TTTT
	55755	TCTTAC	TATCTTTTAGG	AACTTTTTTA	TACAGATCAA	AATTTTGAAT	TAAATACTGT	CTCAATTTAA	AGTC
10	55825	GAGTTC	TTTAAAGTGAA	ATCTCTTCTT	TATAAATGTA	GTGTACTCTA	CCGTACGTAG	CAATACCGTC	ACCT
	55895	TCATCT	CTCTTGATTT	GAAATCTTGG	TGCGTTTATA	TAATCATAAT	AAGCGTCTTG	ATTTTCTCTA	GTGA
	55965	CACCAC	CATATGAAAA	CCTGTGCGCA	TTACGGTTTT	CCGCTTCTTT	AACAACAAAT	ATGTCTAATC	CCGG
	56035	ATTTTT	ACGTGCTTTA	AATCTTTTCAA	TATCTTTACC	AAATATCTGT	ACTCTTGTGA	ATTTTCTATT	TTTT
	56105	TCAAAG	ATAAGGTAAT	GCTTGCCACC	TTTGCTATAA	CGATAACCAG	TAACATTTTT	AAGTTCCTTA	CTTG
15	56175	CGCCAC	TCTAGTAATC	TCTTAAGTCA	AAGATATGTA	TTGTCAACAT	TTTCATATTTT	GCTTTTACTT	CAGT
	56245	CGCATT	TACAGTTTGA	TGCAATGACG	TTATGTGTTCC	TGTTGCTAAA	ATACCTAATG	CTAAACTTGC	TTTC
	56315	GCATAT	GCTGTCAAT	TCATAGTTGT	ATGCTCCATT	CGTAATTATT	AGATTTGTTC	GCTTACGCTT	ATTG
	56385	AATCAT	ACAGCTTTAT	TATAGTTAGC	GTATTTGACC	TTTCACATTA	AACCATGTTT	AATAATCATT	GAAT
	56455	CATTAT	TAAGTAAATT	AAGGAATCTA	TAATGTTTCGT	TAAATAAAAC	TGATCCCGTT	TGCTTTCACA	CCCG
20	56525	ATAGAT	AGGGATTAC	AGATAAATTC	AGGTCTCTTC	CACGTCATAT	TTGGACCCAT	CGAAAAATCG	GGTT
	56595	CTCAAA	TCATCGAACA	TAACAAAAGA	AGCTAAGCAA	CATGTAGGCC	GTGTGCACTT	AACCTCTTGT	TTTT
	56665	CCGATG	ACAGCTTCTA	TTTAGAGAAT	GTCAATATTA	TTTTTATATC	ACTTCAATGT	TATCAATATT	AGTG
	56735	CCATCT	ATGACATCTG	CCATGCGATT	TTCTTGTAAT	TTTTTGTGCA	ATTCAAACGT	GTACTTTCCA	CCGT
	56805	TTTTCA	TTTTAATAAC	AATTTTACCT	GAACCAACGT	TACCGTACAG	ATTATTTTTT	TCAATAAGTT	GTTT
25	56875	TCTCAA	TTTAAATCA	AGTCTTTTCA	AGGAAATCTG	TTCTTTAGTA	ATCTTGAATT	CTGAAACATC	ATGA
	56945	GAGATT	GTTACCTTAT	TATCTTCTCT	AGTAATTCCT	ACTCCTGCTT	TGTGATCAAC	TTTTTTACTA	TTAC
	57015	TCTTTG	TGATACCACC	GACAGAATAT	TTTTCCAGAT	TGTAATTATT	TTCTTCTAAA	ACGACAAATA	CATC
	57085	GACATT	CCTATGTACT	CCTTCACCAT	ATTTTTTATC	ATCTTTACCA	ACTAAAGCAA	TTTTATATAT	GAAA
	57155	TAATCT	GGGACAACAT	TCATAAATCT	TATGTGCTGC	CATTTTTTTA	AAATAAATACC	AATCTCATTT	TTAA
30	57225	ATTCTA	AACCTGGTTT	CGTATAATAC	GCTCTTAAAT	CTTTAAATTT	AGGATTTATT	TCTGTTGGTA	CTTG
	57295	TTTTGT	GGTTGGCGAT	TGTGGTGTGT	CTGATTTAGT	AGATTGCATT	GGTTGTGGCG	TGTTTGTGTA	TGGA
	57365	GGTGT	GTCACCTTAG	TTGAAGGCGG	TGTTGTCGCA	TTTGCTGTTT	GTTCGCGTGT	TTCTACTTTA	GTTG
	57435	AGGGCG	GTTTGTGTCG	GTTTGGTTTT	GATTTGCGGT	CTTCTATTTT	AGTTGAGGGC	GGTGTCTGAT	TGGG
	57505	TGCTTC	CACCTTAGTG	GAAGATAGTG	TTGTCGCGTT	TGCTGCTTGC	GTGTCGTTG	TGATTACACC	TGTT
35	57575	GTTAAA	AGGCCTAGTG	CTAAACTTGT	TTTAGCAATC	GTTGTTATTT	TCATAGTTGT	ATGCTCCATT	CGTA
	57645	ATTATT	AGATTTGTTT	GATTACATTC	ATTGAATCTT	ACAGCTTTAT	TATAGATGGC	TATGCTCTCC	ATTC
	57715	ACATTA	AACCTTGTTT	AACTATATTT	GAATCATCGT	TAAGTAAATT	AAGAAATCCA	TAATGTTTCG	TAAA
	57785	TAAAAA	TGATTTTGAT	GTGATTCAC	ACTTGGCACA	TTTGAAGTTT	CGTCACCTTT	AAGACATAGA	AATG
	57855	CCACTT	TTACAAACAA	ATGAATATTC	GTCTTTTTTAC	ATCATTACGC	ATAATAAAAG	AGCTAAGCA	ACAT
40	57925	GTA AAC	CTTTGTCAC	TAACCTTTTG	TTTTTCCGAT	GACAGCTTCT	ATTTAGAGAA	TGTCATGATT	ATTT
	57995	TATATT	CACCTTCAAT	TTATCAATAT	TAGTGCCATC	TATGACGTCT	GCCATACGAT	GCTCTTGCAG	TTTT
	58065	TTGTGT	AATTCAAACG	TATATTTCCC	ACGTTTCTTC	ATTTTAAATA	CGATTGTTCC	TGAACCCATG	TTAC
	58135	CGTAAA	GATTATGTTT	TTCAATAAGT	TGTTTTCTCA	ATTTAAATAT	AAGCTTTTTT	AAGGAAATCT	CTTC
	58205	CTTAGT	AATCATGTAT	TCTGAAACAT	CGCGTGAAAT	CATACCTTGA	TTATCTTTTT	TAGTAATGCT	TAAT
45	58275	TCTACT	TTGTGATTAA	CTTTTTTACT	ATTAGTCTTC	GTGATGCCAC	CGACAGAATA	TTTTTTTCAAT	TGAT
	58345	ATTTAT	TGTCTTCTAA	AACGATAAAT	ACATCGATAT	TATCGTAAGG	TCCATCTTTA	TATTTTTTCT	CATC
	58415	TTTTCC	AACTAAAGCT	ATTTTATAGA	TGAACCTATT	TGGAATAACA	TTCAATAAAC	TAACCGTCGT	CCAT
	58485	GGTTTG	AGCATAAATC	CAAACTGCTT	TTCAAATTCA	AAACTCGGTT	TTGTATAATA	CGCTCTTAAA	TCTT
	58555	CATATT	TAGGAGTCAT	ATCTGTTTGT	GCTTGTTTTA	TGGTTGGAGA	TTGTGGTGTG	TCTGATTTAG	TAGA
50	58625	TTGCAT	TGGTTGTGGC	GTGTTTGTGT	ATGGAGGTTG	TGTCACCTTA	GTTTTCGGCG	TTGTGGATTG	GGTT
	58695	GTCGTT	TGTGATTGTT	CTTGTTTAA	CGTGGCGGTT	GCTGATATAT	TAAGCGTTTT	CTGCTCTTCT	TGTT
	58765	TAGGTT	GTGATATTTT	TCTATTTTGT	GAAGCTGAGG	TTTTTTCCTC	ATTAGTATTT	GGTGCCTTTT	CGAG
	58835	TTTAGG	CGTGCGTTCT	TGTCTTGTGT	TAGCTGCTTG	TGTTGTGCGT	GAATTTGCAC	CTGCTGTTAT	GTTT
	58905	ATCAT	GCTAATCGCT	CTGCTTTAAG	CGTTGGTACT	TTGTCAACTT	TAGTTGATTG	TATTTTTTCT	GCTT
55	58975	TGACCG	ATTGCGTCGT	TACTGTAATT	GCGCCTGTTG	TTAAAAGCCC	TAGTGCTAAA	CTGGTTTTAG	CAAT
	59045	TGTTCT	CATTTTCATA	ATTGTATGCT	CCAATCTATA	TTATATTTCGA	TTGTCTTTTT	ACGTAATTTG	AATC
	59115	ATACAA	CATCATTATA	GATGGCGTTC	TAAGATAATC	ACATTA AAC	CCTTTTAAAC	ATTATTGAAG	TATT
	59185	ATTAA	TAATTTAAGC	AAAAAATAAT	GAGTGAGTAT	GAGATTAATA	TAGCCTTTCT	ATGTGCCTTT	GAAA
	59255	TAATTT	TTAAGCATT	AAAAGAAATT	AAGCAACGTT	TGATCGTCAC	TTAACCTCTC	TATTTCAATT	TCAA
60	59325	CTTATT	TCGTATCAAA	GTATATGTGT	TATGCTTTTA	TAACCTTTGAT	TTCAATTTCTA	TCAATATCTG	TGAC
	59395	ATTGAT	AACATCGGAC	ATACGGTCTT	CTTGTAACCT	TTTATCCAAT	TCAAATGTAT	ACTTTCCATA	GTAT
	59465	TTCTTT	TTGACTGTAA	TTTTTCTCTG	ACTCATTTCA	CCGTAAAGAC	CATAATTATC	AATAAGGTAT	TTTC
	59535	TTAATT	TAAAAATCAAT	CTCTTTCAAT	GACATCGGTT	CTTTATCTAT	TTTAAATGGG	AAAAAGTCAT	AATC
	59605	ATATTC	ACCAGTATGA	TCTTCTTTAA	TAACCTTTGC	TTCTGCTATT	AGGTCGACAG	CTTTATCGTT	TGCA
65	59675	CTCGTG	ATACCCCCAA	TAGAGTACTT	TGCACCTTCA	AATCTCTTAT	CCTCATTAAC	GTAAAATATA	TTAA
	59745	GATTAC	GATGTACACC	CGTATGATAA	TGTTGCTTAT	CTTTGCCAAT	TAAAGCAATA	TTATTAACAG	AATT
	59815	ACCATC	TATGATATTC	ATAAATTTAA	TACTTGGTTG	AATGAAACTG	ATATAACCTG	TCACATTTTT	ATAT
	59885	TCAATA	CTAGGTTGAT	TATAATAAGC	TTTTAATTTT	TTGCTATTTT	CACCTATTAC	AATAGGTTTC	TTTT
	59955	CGGCAT	GAACTGGTTT	TTCCGTTGTA	GTGTTTACAC	CTGTTGCTAA	TATTCCTAAT	AACAACCTTA	TTTT
70	60025	TGCAAT	ATTTTTTCATT	TTCATAGTTG	TATGCTCCAA	TCTATTATAA	TTAGATTGTT	TTATTACGTA	ATTT
	60095	GAATCA	TACACCCATA	TTATAGGAGC	TGTATTCCGA	TATTCACATT	AACCTGTTTT	TAACATTTAA	TAAA
	60165	ATATGA	TTAAGCTATT	TAAGCAAAAG	ATAATCCGTG	GTGCGACTGT	TTGTGTGAGT	TTTGTAACTT	TTCC
	60235	CCTCAA	AAAAAGAAGC	TAAGCAACAG	TACGTCGCTT	AACCTCCATA	TATTTCCATA	ACAAGCACTA	TTCA
	60305	ACATAG	CGCTTTGTTG	TTAGAAAGTA	TTATTTTCATT	TCTACTAGAA	TTTTATTAAT	CTTAGTGCCG	TGCA

5 60375 TTGACT CACCCATACG TTCTTTTTC AAGTTTTTGAC TTAAATCGAT TGTATGTGTT GTGCCATCAT TCAT  
 60445 TGTAAT TGTAATTTGA CCTTGTTTAA GACCAATTTGA ATACAAGCCG TGTGTTTAA TCGCACGTTT TCTT  
 60515 AATCTA TAATCAAGTT CTTTTAATGA GATGCTTCT TTTGAAATGT AGTAAAAATC TTAAAGTACA TCTT  
 60585 CATTAT CTTTCTTGAT TTCTAAAATT GCGGTATTGA TATAATCAAT ATACTGCACA CTATTTGATT TTGT  
 60655 AACGCC ACCAATTGAG TAATTGTTG CTTGCTACT GTCACCTCT CTAAGGATAA ACACATCTAT ATTT  
 60725 GAGTTT TCTCCGTCTT TAAATTTATC TTTATCAGAT CCAAGTAAAG TGATTCGAGA ATGCTGTTGC CCTA  
 60795 CTGTGA CTTCTAAATA GTGTTTACCC TCATCAGTAT ATTTAAATCC AGTCACATTT TTACGCTCTA AAAT  
 60865 TGGTTT ATTATAATAG TGTTTTAACT CTGATTCACCT TTGTTGTTTA ACTTCTGCTT TCGCTTGATC TGAT  
 10 60935 TGTACA TTCGATGTAA TTACACCTGT TGCTAACATT CCAATGCTA AACTTGCTTT TGCTATCGCT TTAA  
 61005 ATTTCA TAATTTTATG CCCCARTCTA TTTAAATTTG ATTTTCAAT TACGTAATAT GAACAATCTA CCTA  
 61075 CATTAT ATTGAGGCGC TATCCCTTT CAAATTAAAC CTCATTTAAC TAAACTTGAA TCATTGTTAA GGTA  
 61145 ATTAAT AAAATGCTAA CTGTTATTAA TCAATTCTTA TCCGTCGAGA CTCTTTTATA TACAAATCAC ACTT  
 61215 TATACC ATTTTATTTT TAAAACTGT AAAACTTTCA GACACAAAAC AGACATCATA TTGTCGTAGT ACCT  
 15 61285 GATTAA AATTTTCAAT GATATTAGTG ATACACATCA ACTAAAACGC GTCATTACGG CATTTTCAAC ATTA  
 61355 ATAATA TTTACGTATT TACTATACAC GTCGACAATA AATACAGATT AGAATTTTTT ATAAAATCAA ATGT  
 61425 TTTGAA AACCATCAA TAAGTATTAT GTTATGTATA TAAAGATTAG GAGATGAAGT GAATGAATAT TATG  
 61495 TTAACA GGTGCTACAG GTCATTTAGG CACACATATT ACAAATCAAG CCATTGCAAA TCACATAGAT CATT  
 61565 TTCACA TTGGTGTTAG AAATGTTGAG AAAGTCCAG AAGATTGGCG CGGAAAAGTT CCTGTTGAC AATT  
 61635 AGATTA TTTTAATCAA

## SEQ ID No:38 - SSL1

25 gene 60331..61026  
 /gene=" NCTC8325ssl1"  
CDS complement 60331..61026  
 /gene="NCTC8325ssl1"  
 /product="(SET6) SSL1"  
 /translation="MGLKIMKFKALAKASLALGMLATGVITSNVQSVQAKAEVKQOSESELKHYYNKPILERKNVT  
 30 GFKYTDGKHYLEVTVGQQHSRITLLGSDKDKFKDGENSNIDVFILREGDSRQATNYSIGGVTKNSVQYIDYINT  
 PILEIKKDNDVLKDFYISKEDISLKELDYRLRERAIKQHGLYSNGLKQGQITITMNDGTTHTIDLSQLKLEKERM  
 GESIDGTEKINKILVEMK"

## SEQ ID No:39 - SSL2

35 gene 59350..60045  
 /gene=" NCTC8325ssl2"  
CDS complement 59350..60045  
 /gene="NCTC8325ssl2"  
 /product="(SET7) SSL2"  
 40 /translation="MKMKNIKISLLGLATGVNTTEKPVHAEKKPIVISENSKKLKAYYNQPSIEYKNVTGYI  
 SFIQPSIKFMNIIDGNSVNNIALIGKDKQHYHTGVHRNLNIFYVNEDKRFEGAKYSIGGITSANDKAVDLIAEARV  
 IKEDHTGEYDYDFFPFKIDKEAMSLKEIDFKLRKYLDNYGLYGEMSTGKITVKKKYGKYTFELDKKLQEDRMSD  
 VINVTIDIDRIEIKVIKA"

## SEQ ID No:40 - SSL3

45 gene 57989..59059  
 /gene=" NCTC8325ssl3"  
CDS complement 57989..59059  
 /gene="NCTC8325ssl3"  
 /product="(similar to SET8) SSL3"  
 /translation="MKMRTIAKTSALGLLTTGATVTTQSVKAEKIQSTKVDKVPPLKAERLAMINITAGANSAT  
 TQANTRQERTPKLEKAPNTNEEKSASKIEKISQPKQEQKTLNISATPAPKQEQSQTTESTTPKTKVTPPST  
 55 NTPQPMQSTKSDTPQSPTIKQAQTDMPKYEDLRAYYTKPSFEFEKQFGFMLKPWTTVRFMNVIPNRFYKIALVG  
 KDEKKYKDGPDNDIVFVLEDNKYQLKKYSVGGITKTNSSKKNHKKVELSITKKDNQGMISRDVSEYMITKEEISL  
 KELDFKLRKQLIEKHNLGNGMSGTIVIKMKNNGGKYTFELHKKLQEHMADVIDGTNIDNIEVNIK"

## SEQ ID No:41 - SSL4

60 gene 56698..57624  
 /gene=" NCTC8325ssl4"  
CDS complement 56698..57624  
 /gene="NCTC8325ssl4"  
 /product="(SET9) SSL4"  
 65 /translation="MKITTIKTSALGLLTTGVITTTTQAAANATLSSTKVEAPQSTPPSTKIEAPQSKPNATTP  
 PSTKVEAPQQTANATTPSTKVTPPSTNTPQPMQSTKSDTPQSPTTKQVPTEINPKFKDLRAYYTKPSLEFKNEI

GIILKKWTTIRFMNVVPDYFIYKIALVGKDDKKYEGGVHRNVDVFFVLEENNYNLEKYSVGGITKSNSKKVDHKAG  
VRITKEDNKGITISHDVSEFKITKEQISLKEIDFKLRKQLIEKNNLYGNVSGSKIVIKMKNNGGKYTFELHKKLQENR  
MADVIDGTNI  
DNIEVNIK"

5

SEQ ID No:42 - SSL5

10        gene                    55630..56334  
                              /gene=" NCTC8325ss15"  
                              complement 55630..56334  
                              /gene="NCTC8325ss15"  
                              /product="(SET10) SSL5"  
15        /translation="MKMTAIAKASLALGILATGTITSLHQTVNASEHKAKYENVTKDIFDLRDYYSGASKELKNVT  
GYRYSKGGKHYLIFDKNRKFTRVQIFGKDIERFKARKNPGLDIFVVKAEENRNGTVFSYGGVTKKNQDAYDYINA  
PRFQIKRDEGDGIATYGRVHYIYKEEISLKEIDFKLRQYLIQNFDLYKKFPKDSKIKVIMKDGGYYTFELNKKLQT  
NRMSDVIDGRNIEKIEANIR"

20        SEQ ID No:43 - SSL6

gene                    54489..55184  
                              /gene=" NCTC8325ss16"  
                              complement 54489..55184  
25        CDS                    /gene="NCTC8325ss16"  
                              /product="(SET8) SSL6"  
/translation="MKLKTLAKATLVLGLLATGVITTESQTVKAAESTQGQHNYKSLKYYYYSKPSIELKNLDGLYR  
QKVTDKGVYVWKDRKDYFVGLLGKDIEKYPQGEHDKQDAFLVIEEETVNGRQYSIGGLSKTNSKEFSKEVDVKVTR  
KIDESSEKSKDSKFKITKEEISLKEIDFKLRKKLMEEKLYGAVNNRKGKIVVKMEDDKFYTFELTKKLQPHRMGD  
30        TIDGTIKIEINVELEYK"

SEQ ID No:44 - SSL7

35        gene                    53373...54068  
                              /gene=" NCTC8325ss17"  
                              complement 53373...54068  
                              /gene="NCTC8325ss17"  
                              /product="(SET1-C) SSL7"  
40        /translation="MKLKTLAKATLALGLLTGVTITSEGOAVQAKEKQERVQHLYDIKDLHRYYSSESFEFSNISG  
KVENYNGSNVVRFNQENQNHQLFLLGKDKEKYKEGKDVFFVKELIDPNGRRLSTVGGVTKKNKSSETNTHLFV  
NKVYGGNLDASIDSFSINKEEVSLKELDFKIRQHLVKNYGLYKGTTKYGKITINLKDGEKQEIDLGDQLQFERMGD  
VLNSKDINKIEVTLKQI"

45        SEQ ID No:45 - SSL8

gene                    52331...53029  
                              /gene=" NCTC8325ss18"  
                              complement 52331...53029  
50        CDS                    /gene="NCTC8325ss18"  
                              /product="(SET12) SSL8"  
/translation="MKFTVIAKAIFILGILTTSMITENQSVNAKGKYEKMNRLYDTNKLHQYYSGPSYELTNVSG  
QSQGYDSNVLLFNQQNQKFQVFLGKDENKYKEKTHGLDVFAVPELVLDLGRIFSVSGVTKKNVKSIFESLRTPN  
LLVKKIDDKDGFSIDFFFIQKEEVSLKELDFKIRKLLIKKYKLYEGSADKGRIVINMKDENKYEIDLSDKLDFER  
55        MADVINSEQIKNIEVNLK"

SEQ ID No:46 - SSL9

60        gene                    51253...51951  
                              /gene=" NCTC8325ss19"  
                              complement 51253...51951  
                              /gene="NCTC8325ss19"  
65        /product="(SET13) SSL9"

/translation="MKLTTHAKATLALGILTTGVFTAESQTGHAKVELDETQRKYYINMLHQYYSEESFEPTNISV  
KSEDDYGSNVLNFKQRNKAQKVFLLGDDKNKYKEKTHGLDVFAVPELIDIKGGIYSVGGITKKNVRSVFGFVSNPS  
LQVKKVDAKNGFSINELFFIQKEEVSLELDFKIRKLLIEKYRLYKGTSDKGRIVINMKDEKKHEIDLSEKLSFER  
MFDVMSDKQIKNIEVNLN"

SEQ ID No:47 - SSL10

gene 50204...50887  
/gene=" NCTC8325ssl10"  
CDS complement 50204...50887  
/gene="NCTC8325ssl10"  
/product="(SET14) SSL10"

/translation="MKFTALAKATLALGILTTGTLTTEVHSGHAKQKSVNKHDKALYRYTGTGKTMEMKNISAL  
KHGKNNLRFKFRGIKIQVLLPGNDKSKFQORSYEGLDVFFVQEKRDKHDIFYTVGGVIQNNKTSVGVVSAPILNISK  
EKGEDAFVKGYPYIKKEKITLKELDYKLRKHLIEKYGLYKTISKDGRVKISLKDGSFYNLDLRSKLKFKYMGVEI  
ESKQIKDIEVNLK"

SEQ ID No:48 - SSL11

gene 46120...46797  
/gene=" NCTC8325ssl11"  
CDS complement 46120...46797  
/gene="NCTC8325ssl11"  
/product="(SET15) SSL11"

/translation="MKLKNIKASLALGILTTGMITTTAQPVKASTLEVRSQATQDLSEYYNRPFFFEYTNQSGYKE  
EGKVTFTPNYQLIDVTLTGNEKQNFEDISNVDIFVVRENSDRSGNTASIGGITKTNGSNYIDKVKDVNLIITKNI  
DSVTSTSTSTYTINKEEISLKELDKLRKHLIDKHNLYKTEPKDSKIRITMKDGGFYTFELNKKLQTHRMGDVID  
GRNIEKIEVNL"

SEQ ID No: 49

1047303 TATGAAAA AGAACATCAT GAATAAATTA GTTTTATCAA CAGCATGTGT ACTTTTAGAA ACTACATCAA CA  
1047373 CAACCTCC TAAACACCA ATCAGTTTTT CATCTGAAGC AAAAGCCTAT AATATCAGTG AAAACGAGAC TA  
1047443 ATATCAAT GAACTAATCA AATATTACAC TCAGCCGCAT TTTTCATTAT CTGGAAAATG GTTATGGCAA AA  
1047513 GCCCAATG GTAGCATTCA TGCAACATTG CAAACGTGGG TTTGGTATAG TCATATTCAA GTGTTTGGAT CC  
1047583 GAGAGTTG GGGAAACATT AATCAGTTAA GAAATAAATA CGTTGATATA TTTGGAACTA AAGATGAGGA CA  
1047653 CAGTTGAA GGTACTGGA CTTATGATGA AACATTTACT GGTGGTGTTA CGCCAGCAGC TACTTCATCT GA  
1047723 TAAGCCTT ATAGACTATT TTTAAATAT AGTGATAAAC AACAACTAT CATCGGTGGA CATGAATTTT AC  
1047793 AAAGGAAA TAAACAGTA TTAACCTTAA AAGAATTAGA TTTCCGTATT CGTCAAACAT TAATAAAAAA TA  
1047863 AAAAGTTA TATAACGGAG AATTTAATAA AGGTCAAAT AAGATAACTG CTGATGGAAA TAATTACACG AT  
1047933 TGATTATA GTAAAAAGTT AAAATTAAC GACACAAACC GTTATGTTAA AAATCCTCGT AATGCAGAAA TT  
1048003 GAAGTCAT ACTCGAAAAA TCTAACTAAC CTATTACCTT TTGTAATGCG GGATAATTTT AATTATCTAA TT  
1048073 AACCCCTT TTTATAATTA AACATTCCAA CAATACTCAA AGGAGAAAT CGAATGAACA ATAACATCAC GA  
1048143 AAAAAATT ATTTTATCAA CAACATTGTT ACTATTAGGT ACAGCATCTA CACAATTTCC TAATACACCT AT  
1048213 CAATTCTT CATCTGAAGC GAAAGCTTAT TATATAAATC AAAACGAAAC TAACGTTAAT GAGTTAACTA AA  
1048283 TATTACTC GCAAAAAATAT TTAACCTTCT CTACAGTAC GTTATGGCAA AAAGATAACG GTACGATTCA TG  
1048353 CAACGTTG TTACAGTTT CTTGGTATAG TCATATTCAA GTTTATGGAC CTGAAAGTTG GGGCAATATC AA  
1048423 CCAATTAA GAAATAAAG CGTTGATATT TTTGGCATAA AAGACCAAGA AACCATTGAT TCTTTTGCAT TA  
1048493 TCTCAAGA AACGTTTACT GGTGGTGTTA CTCCTGCAGC AACATCTAAC GATAAACACT ATAAACTGAA TG  
1048563 TAACATAT AAAGATAAAG CAGAAACGTT TACTGGCGGA TTTCCAGTTT ATGAAGGCAA TAAGCCTGTT TT  
1048633 AACTTTAA AAGAATTAGA TTTCCGTATT CGTCAAACAT TAATTAAGG TAAAAATTA TATAATAATT CT  
1048703 TATAATAA AGGACAAAT AAAATAACAG GTGCAGACAA TAACTACACA ATAGATTTAA GTAAAGGTT GC  
1048773 CATCAACT GATGCAATA GATATGTTAA AAAACCTCAA AATGCAAAAA TTGAAGTTAT CCTCGAAAAA TC  
1048843 AAATAAC AATAATAATG GAGTTAATAA AAATAATCGC AAATACTATA TTGACTTCGC TCACATTGAA AT  
1048913 TTCTTATT CCTCGTATCA TGATTCTCT GAAAGGAGAT GTTCTAATGA GTAGAACAT CACGAAAAAT AT  
1048983 AATTTTAA CGACAACATT ATTACTATTA GGTACTGTAT TACCTCAAAA TCAAAAACCA GTATTTAGTT TT  
1049053 TACTCTGA AGCTAAAGCT TATAGCATTG GCTAAGATGA AACTAACATC AATGAATTA TTAATATTA CA  
1049123 CACAGCCT CATTTTTCAT TTTCAATAA ATGGCTATAT CAATATGATA ATGGAACAT TTATGTTGAA CT  
1049193 TAAGAGAT ATTCATGGTC AGCACATATA TCTTTTGGG GCGCTGAAAG TTGGGGAAT ATTAATCAGT TA  
1049263 AAAGATCG TTACGTAGAT GTGTTTGGAC TAAAGACAA AGATACTGAT CAGTTATGGT GGTCTTATAG AG  
1049333 AGACATTT ACAGGTGGCG TTACACCAGC GCAAAACCT TCTGATAAAA CTTATAATCT TTTTGTGCAA TA  
1049403 CAAAGATA AACTACAAAC GATTATTGGT GCGCATAAAA TATACCAAGG CAATAAACCA GTATTAACAT TG  
1049473 AAAGAAAT CGATTTCCT GCACGAGAAG CGTTAATAAA AAATAAATA TTATATAACG AAAATCGTAA TA  
1049543 AAGGTAAG CTTAAGATCA CCGGTGGCG TAATAACTAC ACTATTGATT TAAGCAAAAG ATTACATTCA GA

1049613 TCTAGCAA ATGTTTATGT TAAAAATCCT AATAAAATAA CTGTTGACGT CCTCTTTGAT TAGTATATGA AG  
1049683 G

5

SEQ ID No:50 - SSL12

10        gene                    1047304.. 1048029  
             /gene=" NCTC8325ss12"  
             CDS                    1047304.. 1048029  
             /gene="NCTC8325ss12"  
             /product="(similar to SA1011)SSL12"  
15 /translation="MKKNIMNKLVLSTALLLLETTSTQLPKTPISFSSEAKAYNISENETNINELIKYYTQPHF  
SLSGKWLWQKPNNGSIHATLQTVVWYSHIQVFGSESWGNINQLRNKYVDIFGTDKEDTVEGYWTYDETFTGGVTPAA  
TSSDKPYRLFLKYSKQQTIIIGGHEFYKGNKPVLTLLKELDFRIRQTLIKNKKLYNGEFNKGQIKITADGNNTIDIL  
SKKLKLTDTNRYVKNPNRAEIEVILEKSN"

20        SEQ ID No:51 - SSL13

gene                    1048124.. 1048849  
             /gene=" NCTC8325ss13"  
             CDS                    1048124.. 1048849  
25        /gene="NCTC8325ss13"  
             /product="(similar to SA1010 from strain N315)SSL13"  
/translation="MNNNITKKIILSTLLLLLGTASTQFPNTPINSSSEAKAYYINQNETNVNELTKYYSQKYL  
TFSNSTLWQKDNGTIHATLLQFSWYSHIQVYGPEWSGNINQLRNKSVDIFGIKDQETIDSFALSQETFTGGVTPAA  
TSNDKHYKLVNVTYKDKAETFTGGFPVYEGNKPVLTLKELDFRIRQTLIKSKKLYNNSYNKGQIKITGADNNYTIDIL  
30 SKRLPSTDANRYVKKPQNAKIEVILEKSN"

SEQ ID No:52 - SSL14

35        gene                    1048957.. 1049673  
             /gene=" NCTC8325ss14"  
             CDS                    1048957.. 1049673  
             /gene="NCTC8325ss14"  
             /product="(similar to SA1009 from strain N315)SSL14"  
40 /translation="MSKNITKNIILTTLLLLLGTVLPQNQKPVFSFYSEAKAYSIGQDETNNINELIKYYTQPHF  
SFSNKWLYQYDNGNIYVELKRYSWSAHISLWGAESWGNINQLKDRYVDVFGCLKDKDQTDQLWWSYRETFTGGVTPAA  
KPSDKTYNLFVQYKDKLQTIIGAHKIYQGNKPVLTLLKEIDFRAREALIKNKILYNENRNKGLKITGGGNNYTIDIL  
SKRLHSDLANVYVKNPNKITVDVLF"

45

***S. aureus* strain EMRSA 16(252) taken from unpublished  
genome project at the Sanger centre via  
<http://pedant.gsf.de>**

50        SEQ ID No:53

122949 GC CAGCTTTATC AGTTGAAGCT TCTTCCACCT CTTCATCTTC TTCATCTTCA CCTAAATCAG GGCTTGAC  
123019 GA CGAAACCTCA ATTTCTTGAT TATGAACAAT TTTACGATTA TTTAAAGGGT ACTCAAATTC TCCTTTGT  
123089 CA TTTGCTTCAA CAAAACCCAA ACCGCCATTT TCTACTGAAT CCGCACTTTT CCCATCAACT GTTAATAA  
55 123159 AA CATAGTGATT TGGTAAAGTA GTTCCCGTTA TCTTTTGAGC TCCAGGTTTA ATAGTATCTA ATTTAACA  
123229 AA GTTCTTATTA GATTCTACTG TTTTGTGTTT TTTTGAATTT TGTTCAGTCT GTGAACAAC TTTTGATT  
123299 CT TGATTGTGAC TTGGCTTCTC ATTAGCCTCT GATGCTTCTG CTGTCCGTT TGTCATGATA TATAACAT  
123369 TG TAATTGCAAC AGATACTAAC CCGACTTTCA TTTTACGTAA CTTAAATTT TCCCTCATGA TATACTCC  
123439 CT CGAATATTAA TATAAATCGA CTTCATTTT TTCTATATTT CTACCATCAA TTACATCACC CATACGAT  
60 123509 GA GGCTGTAATT TTTTATTTAA TTCAAAAGTA TAGTAGCCGC CACCTTTTCAT AGTAATTTCTA ATTTTACC  
123579 GT CTTTAGGCTC TGTCTTATAA AGTTCATGAT TTTCAATTAA ATGTTTCTT AATTTGAAAT CTAATCTC  
123649 TT TAATGAAATT TCTTCTTTAT TAATTCTATA AGTTTCTGCT TGTCTTGTTG CTGTATTATG TCCTGTTG  
123719 GT TTAGTTATTT CTAGACCTAC ATTATTTACG AAATCTTTAT AATCGTTTTT ATTCGCTTTG GTAATTC  
123789 AC CTATTGAAAT ATTTTCAGCT TGTTTACCTG ATCCTTCTCT GACTACAAAG ACATCTAACC CTTTCATAA  
65 123859 TC ATAGTCATAA TCTTTAAATC TTTCTTTGTC TGTACCAAGT AAAGTAACTA CATTAAGTTG TGTCCCAT  
123929 CA ATAATGTTCA TTTTATCCTT TTCTCTATAA CCACTCACAT TTTGAAAATT ATATCCTGTT CCACTGTA

123999 GT ATTTTTTTTAA TTCTTGCGTG TCGTTTGAAG TAACTGATAA TCTGCTTTGC TCACTTGCTT TTA CTGCTTGGC  
 124069 TG AGCAGTAGTT GTAATCATCC CTGTTGTCAA AATCCCTAGT GCTAAACTTG CTTTAGCAAT ATTTTFTA  
 124139 AT TTCATAATTC TATGCTCCCA ATTTTGTAGC TATTTGATT ATTCTATTAC GCAATACGAA CAATCCTC  
 124209 AT TCATTATAAT GAGCAACGGT CGTGCTTCAC ATTAACTTA CTTTAACTAA AAATTAATCA TTATTAAAG  
 124279 TG ATTTAATAAA TAGTTAACTA ATATCATCA ATCCCTATCG ATCTACTCTC TTTTATTTAC GAATAACA  
 124349 CT TTATCTCAAC TTAATCTTTA TTTAATCTACT TATTTCTGTAC ACAATTTGCGA CACAAAAAAG ACACGTG  
 124419 CT ATCACAGTGC CTAAATTTAA TGTTCGTTTT TAATTTACTT TATGTTTATA ACTTTTATGT TAAATTAC  
 124489 AC AAATTTTATT CTATGCAAC CTATGTGTTT ATCAAAGTTA TAAGAACATC TTTTGTAAAA AGGATTGT  
 124559 TT CTCTTTTTCT AATATTTCAA TTTTATTTT CTGTTTACTA ATCAATATAT CCATTTTTTT AAAGAAAT  
 124629 CA CCTATCTTTT CTGTTCCTC CAATACAGGT ATATCTATAT TTATATTTTT TAATTTGTTA TATTTTAA  
 124699 GT TCCATGTATC TGATGTTAAT CCTGTGAAT TAATTTTAA TTTATGAATC ATTCTATGTG TTTTAAAC  
 124769 TT ATATCCAATA AATAATGAGC TAGTATTTTG TGTGGGATA AGCACAGTAT ATGCGGGCT AACAATCC  
 124839 CA TTATAATTTG ATCTACCACT AGCCCTTGCG CACATTCTCA TAGAATTATA TGCAATATCA TTTTTCCT  
 124909 AA CTACTTTATA ATTACTTTT TCTTACTTTG AATATCTATT TCTATCCAAT TCTTCAAAAT TTATATAG  
 124979 CC ACTATTTATA GTTACTGAAA GCATTTGACC TTTGTGAGAA CGTTCGTTTC TCTCTTTTAA ATATTTTT  
 125049 CT ATTTTGTCTAT TTTCCCAATG TGGATAATCT TCACTATTCT CATCTTTGAA TCGCAGTTCC TGTGAGAA  
 125119 GA TTTTCTGCAT ATAGCCTTT TTTTGTGTTT GAAGTAATC AAGCTTTTGT TCTTCTAATT CAATTGTG  
 125189 CG GTCGAGTTTG CTGAAGAAGT CACCTATTTT TCTCTGTTCT TTAGCCGAAA CAGGGTATAT GACCTTCA  
 125259 TT TTATTAATTG CGCTACCTGA TGTAAAGCT CTAGTTGTCA TAGAACTTTT TGTAAATCAT TCTTTTCT  
 125329 AA ATGAATTAGT AAAAAAGACA TATCTTTTAA AATTTATTAT TATTAATCA ATTCCTGAT TAGGCCGC  
 125399 CC TCTTAATACA AATCCACTAA ACACAGGTCA TTTCAAGGTA TTTAAATTA CAGACGGATA ACCTATTT  
 125469 CA CCAATTACCT CACTAGTCTT TGTAAAAAAA ACATCACCT TTTCAACAGA ATAATTTTT AGTTCTTT  
 125539 GC TATTACCATT AACTTTTCCA GTGAGATTA TTTGATTTAA GCTCCTGTTA TTAATACAT CTTTGAAG  
 125609 TT AACCAATCGAC GATCCTGAGC AAAAATATC TTTTCTTTA TTTAAACCAT TTTTAAATTC TAATAACT  
 125679 CG CCAACCTTCT TCTCTTCCCA TTCGCTTCA AACCTGGGA ATCTCAACTC TGGCACATTT TCGGTTG  
 125749 TG TATTACTCAT CTTTCAACAC CCAAGTTCT TTCAGGTATG CATGATTTT TTTGTTCAATT TCTGCGAT  
 125819 TT CTTGTGCGAT ATTTTCAA TCTTGTGGA CTGATCTAA ATCAATTGGC GCTTCTTCTT CGAATGTA  
 125889 TC GACATCTCTC GGTATGTTA AGTTGTAATC GTTATCGGCG ATCTCTTGA ATGTCGCGCT TAGCTAT  
 125959 AC TTATCAATTG TTTCTTACG CTTATATGTC TCTATAATC GTTCGACTTG GGCATCGCTT AAATGGTT  
 126029 TT GGTTTTTTCC TTTTCAA AAA TCATTGGATG CATCGATAA TAATACGTT TCGCTTTGTT GCGCAT  
 126099 TT TTTAAATACT AAGATACATG TTGGAATACT TGTCCTATAG AAAATGTTTCG TGGCAAAAC AATCACGG  
 126169 CT TCTAAGTAGT TCTTTTCTTC AATTAATTA GACGCAATGA TACCTTCTGC GGCACACCG AATAAGAC  
 126239 AC CATGTGGGAG TACAACGGCC ATGGTACCTT CATCATCTAG GTAATGTACC ATGTGTTGAA TAAAGGCA  
 126309 AA GTCTGCTTTA GACTTTGGCG CAAGCTTGCC GTATCCGCTG AAGCGTTCTG CATTTTCAA TTTTGAGT  
 126379 CT GCTGTCCATT TCGCGCTATA TGGTGGGTT GCAATAACCG CATCAAATGT ATGTCTTAAA AAGGCTGG  
 126449 AT TTTCCAACTG ATCATCATTA CGGATATCGA AATTTTCATA TCGCACATCA TGTAAACA TGTTCATG  
 126519 CG CGCTAAGTTG TATGTGGTAT TGTACGTTT TTTGCTCAAAA TAACGATACA CTTTTCGCTC TTTACCAA  
 126589 CG CGTAACAGTA ATGAACCTGA ACCACATGTT GGGTCATATA CATGACGTAA TTTATCTTT CCGTCTGT  
 126659 GA CAATCTTCGC CAGTATCTTA GTACTTTGTT TGGTGTATA GAATCTCGCT GCTTTTTCAC CCGTGTG  
 126729 GC CGCAAAGCGC CCGATAAGAA ATTCGTATGC ATCACCCTAC ATATCAATTT CCATATCACT GTGAACGA  
 126799 AT GGTAAGTCAT CAAGGTTAAT CATAACTTTG GAAATTAACG CAGTACGTTT TTTGACATT TTACCTAA  
 126869 AC GCGTTGAAC TAACTCCATA TCGCTGAACA GTCGATATA GTCATTTTCA TTTTCTTAC CTAGCTGT  
 126939 GA AGTTTCAACT TTACGAATTG CCGTCGCCAG ATGTTTCGATA TCGAAATCTT GCGTTTCAAT TTCACGAA  
 127009 TC ATCGCACTGA ATAAATCTTG TGGCTCAATG AAGTAACCAA CTTGATCAAT TAATCTGCT TTTAAGTC  
 127079 TT CACGGTATTC TTCACTCTGCC CATGCTTCTT GATATGTAAT ATCTTCACTT GCCAAGGCAT CTGCATAT  
 127149 TC TGTGTTCCGCT TTTTCAGATA AGAAGCGATA GAAATCAAG CCTAAATATG AATTACGGAA TTTCACTG  
 127219 CA TCCATATTC CTCTTAAATC ATTGCAATT GACCATAATT TTTATGTAA TTTAGCTTGT TGTGACG  
 127289 TT GTTTTTCAGT AATAGACATG TGATTTCTCC GCTTTTGATG AAGTAATTTG TCTCTTTGTT TAATAGAT  
 127359 TT ATTATAACAT TTGGTGGTGT CGCGATGTTG AGAATTTTGA TGTGATGTTT GGAATTTTAT TATCATG  
 127429 GA GATTGGGATA GGATTCATAA CACGCAACA AAAACGGCTT AACGACGATA TCGATTGATG CATCTTTT  
 127499 TG GTAAGGTATA CGCACCAATT TACTTTTCAT CGTTAAGCCG CTTCATTTAC TTTTATTAT TCATATG  
 127569 AT TTACTTTTAA TTTACTTCAA TATCCTTAAT TTTGTTTACT TCTATGACTT CCCCCATATG TTTGAATT  
 127639 TT AATTTAGTTT TTAATCAAG GTTATAAAG CTACCATCTT TCAGACTAAT TTTAATCCTA CCATCTTT  
 127709 TG AGAGTGTTTT ATAAAGACCG TATTTTTCAA TTAGATGCTT TCTCACTTA AAATCTAAT CTTTCAAT  
 127779 GA TATTTCTTCT TTTTAAATAT CATAGGGTA ACCTTTCACA AAAGCATCTT CACCTTTTTC TTTTGTAA  
 127849 CA TTTAATCTTG GTGTGCTGAC AAAGCCGGAT GTTTTATTGG TCTTTGTTAC ACCACCAACA GTATATGA  
 127919 TA TGTGCTGCTT GTCTCTTCTT TCTTGAACAA AAAACACATC TAAGCCCGTA TGTCTTCGCT GTTGATAT  
 127989 TT ACGGTACTCA TCTCCAGGCA ATAATACCTG AGTCTTCATC CCTCTATAT TAAACGTAAT GTTATTTT  
 128059 TA CCATGTCTCA AAGCATTAAT GTTTTTCAT TCTTAAAGT TTCCAGTGA TATCTGGTGT AATGCTTC  
 128129 TT TGTGATGTTT GTTTACTGAC TTTTGTATTT GTTTAGCATG ACCTGAATGG GCTCTGTGTT TTAAGTT  
 128199 CC TGTGTTTAAA ATCCCTAATG CTAATGTTAC TTTTGCTAAT GCTGTCAAT TCATATTTAT TTAACCTCA  
 128269 AT CTTAATGTAT TGGATTGTTA TTATTACGTA ATTTGAATCA TATACGCTTA TTATAACGTC CATCTTTT  
 128339 AA CGCGATAATT AACCTCGATT TAAATATGCG TGAATATTA TTAATTA AAA ACAATATAT TTTAATGT  
 128409 CC AGGTTAATTA ACTTTACGCG TTTACATGTC GTGTTTTTC CGGATCATCG TTTTCTACAA CTAATGTG  
 128479 GT TATTTCTACT CTTTGAACG CAAAAACAA CTTAACAACA ATAAATGATT TAACACAGTT CTTATTCA  
 128549 AA GGTACGGGTG TCAATGCACA TTTTATCGTT AAGCCGCTTT TTAATTTTTT ATACTATCAT CTTTAAAT  
 128619 AA TTCAAATTTCA CTCAATATT TTTAATCTGC TTTACTATCA GCACATCAAA ATACGATCA AAACCTTAA  
 128689 TT TTTCACTTAA ATCAATTTCA TGTTCCTTTT CGTCTTTCAT ATTGATAACA ATTCTACCTT TATCTGAC  
 128759 GC GCCTTTATAC AATCTATATT TTTGACATA CATTTTTCTG ATTTTAAAT CAAGTTCTTT CAATGATA  
 128829 CT TCTTCTTTTT GAATAAAGAA CAACCTTTT ATCGAAAAGC CATCTTTAGG ATCAACTTTT TTAACCTG  
 128899 CA TCCCGGATG ACTTACATAG CCAACACGG ATCTCACATT TTTCTTTGTT ATACCGCAA CGCTATAT  
 128969 AT GCCGCTTTA GTATCTATTA ATTCAGGTAC TGCAAAGACA TCACGGCCAT GCGTCAGTTT TTTATATT  
 129039 TA TTTCTATCGT CGCCAATTAG AAATACTTTG AAATTTTTAT TTCGTTGGTT AAAGTTTAAA ACGTTAGA

	129109	CC	CATAATAATC	TTCACCTTTA	ACACTAATGT	TTGTTGATTC	ATAGCTTTCT	TGAGAATAGT	AATCTTTT
	129179	AG	CATATTTATA	TAATATTTGC	GTTGTGTTTC	ATCCAACCTT	ACTTTTCGCT	TTACAGTTTG	ACTTTCTG
	129249	CT	GTCATCACAC	CTGTTGTAA	TATTCCTAAT	GCTAATGTTG	CCTTAGCTAT	TGCTGTTAAT	TTCATAAT
5	129319	TT	ATTGCTCCAA	TCTTAATGTA	TTGGATTGTT	ATGATTACGT	AATTTGAAAC	ATCTGCCTTT	ATTATAAA
	129389	GT	GCATCGAAAC	AAACTGACAT	TAACGTGCGT	TTAAATAACT	TTGAATCATT	GTTAATGAAC	ACCTAATT
	129459	AA	TATTCAAAGT	ACGTGGTTTT	AATGAGTGGT	ATGACTTATT	TAGTATACTC	GTGACTTGTG	TTATCAAA
	129529	TG	TAAAAAACAG	CCTAACGACG	ACGTTGCTTT	AAATACGCTC	CCTTGATAGG	ATATACATAT	TTATTCAA
	129599	CG	TTTCATCGTT	AAGCTGCCTC	AAAATTTATT	ACTTTAAAGT	CATTGATTGC	TTACTTTAAA	TTTGGTTA
10	129669	AT	AGTGACTGAT	ATACCTCTAA	TGCTCTTACT	ATTCAACACA	TCGCCCATGC	GCTCGAATTG	TAATTTAT
	129739	CA	CCTAAATCAA	TTTCTACTTT	ATTTTCGTCT	TTCAAATTGA	TAATGATTTT	ACCGTATTTA	GATGTACC
	129809	TT	TATATAATCC	GTAATTATTA	ACTAATTGTT	GTCTAATTTT	GAAATCAAGC	TCTTTTAAATG	AGATTTCT
	129879	TC	TTTTTGGATT	AAAAATGAGT	CAATTGATGC	ATCTAAATCT	TCACCATTAA	CTTTATTAAC	AAATAAAG
	129949	GT	GTATTAGTTT	CAGAAGTTT	GTTGTTTTCT	TTGCTTACAC	CACCAACAGT	AGATAGTCTG	CCGTTTGG
15	130019	AT	CAATTAAATC	TTGTACTACA	AAGACATTTT	GGCCTTGTAG	ACCTTCTTTA	TATTTCTCTT	TATCTTTT
	130089	CC	TAATAAGAAT	AATTGGTGAT	TTTGATCTTT	TGGGTTAAAG	CGTACAACGT	TAGAACCATT	GTAGTTTT
	130159	CA	ACCTTACCAC	TAACATTACT	ATATTGCGAA	CCTTCTGATG	AGTAGTATCG	ATGTAAATCT	CTAATATC
	130229	AT	GTAATGTTG	TACTCTCTCT	TGTTTTCTCT	CCGCTTGAAC	TGCTTGACCT	TCTGATGTAA	TGACACCA
	130299	GT	AGTTAATAAA	CCTAATGCTA	ATGTTGCTTT	AGCTAACGTT	TTAATTTTCA	TAGATTATT	CTCCCAAT
20	130369	CT	ATTTTAAAT	TTTGACTTAA	TATATGTTTA	TGTGATTAAT	TCAATTACGT	AATACGAACA	ATCTACTA
	130439	TC	ATTATATAAG	CCGGATACTG	CGCTATGAAT	TAACCGCTTT	TTAACTATTC	CTGAACCAAT	GTTAAGCT
	130509	AA	TTAATGGAAT	CCTAACAGTA	TTAATCATTT	TTATAAATCT	TATGCTATTT	TTTTCACCAT	TCACACTT
	130579	TC	TTTCAATTTA	ATTATGTTTT	CCACATTTGT	TCATGTCAC	AGTGCACAC	CGTGCAGCTA	TAGGTATC
	130649	AA	CTATTTTCAC	AGTTTTTATT	GTCGTTATTA	TCTTTCATCG	ACCCACTTCA	TTTCCGATCG	TAACGACA
25	130719	CA	AAAAAGAAGC	TAAGCAACCT	ATGCTGCCTA	GCTCATCTAT	ACTATCCATA	TTTTACTATT	ATCCATGT
	130789	TT	CGTTGAATTA	TCTAATATTG	GCTTCTATTT	TCTCAATATT	TCTGCCATCT	ATGACCTCAC	CTATTTCGG
	130859	TT	TGTTTGTAA	TTTTTATTGA	GTTCAAACGT	ATAATAGCCG	CCATCTTTCA	TTGTCACCTT	TATCTPGC
	130929	TA	GCTTTAGGAA	ACTTTTATA	CAGATCAAAA	TCTTGAATTA	AATACTGTCT	TAATTTAAAT	TCAAGTTC
	130999	TT	TCAGCGAAAT	CTCTTCCTTA	TAAATGTAGT	ACTTTTAAAC	ATGCACAGAA	ACACCTGCAC	CTACTTCT
30	131069	TT	TTTAATAACG	AATCTAGGTG	CGCTTAAAGT	ATCATAGTAA	GCGCCTTGAT	TTTTCTTAGT	AACGCCAC
	131139	CA	TATGAATACA	CTGTGCCGTT	GCGATTTTCC	GCTTCTTTAA	CTACAAATAT	ATCTAATCCT	GGATTTTT
	131209	AC	GTTTTTAAAT	CCTTCTATA	TCTTTACCAA	AAATTTGTAT	TCTTGTGAAC	TTTCTATTTT	TATCAAAG
	131279	AT	AAGTGAAGTG	TTGCCACCTT	TGCTATAACG	ATAGCCAGTA	ACATTTTAA	GTTCTTACT	TGCCGAC
	131349	TA	TAGTAATCTC	TCAAGTCGAA	GATATCTTTT	GTCACATTTT	CATATTTTGA	TTTATGTTCT	CTCGCATT
35	131419	TA	CAGTTTGAGC	CGTTGATGTT	ATAACCCGAC	TCGCTAAAAT	ACTTAGAGCT	AAACTGGCTT	TCGCAATT
	131489	GC	TGTCATCTTC	ATAGTTGTAT	GCTCCAATCG	ATTTTATTAG	ATTTGTTTGA	TTACGTTTAT	TGAATCAT
	131559	AC	AGCTTTATTA	TAGTTGGCGT	ATGCTCTCTAT	TCATATTAAA	CCTTGTGTTAA	CTATATTGTA	ATCATCGT
	131629	TA	ATTAAATTTA	GGAATCCATA	ATGTTTCGTTA	AATAAAATG	ACTTTTGTG	TACTTCAACG	CTTGATAG
	131699	AT	TGGTATTTTT	AGACACATTT	AGGACATAAG	CATGTTTATG	GTGGTCAACA	TCAAATATTC	GTCTTCTT
40	131769	AT	ACCACGTGTC	AACACGAAAG	AAGCTGAGCA	ACATGTAAAT	TTTTGTCACT	CAACTTCTCG	GATCATCG
	131839	TT	CAACGGTTCT	ATGCAGTATC	CATCACGATG	ATTTTAGATT	CACCTTCAAT	CTTTCGATAC	TAGTGCCA
	131909	TC	TATGACATCT	GCCATGCGAT	GTTCTTGTA	TTTTTTATGC	AATTCAAACG	TGTACTTCCC	ACCGTTTT
	131979	TC	GTTTTAATAA	CAATTGTTCC	TGAACCAATA	TTACCGTACA	AATTATGTTG	TTCAATGAGT	TGCTTTCT
	132049	CA	ATTTAAAAATC	AAGCTCTTTC	AAGGAATACT	CTTCTTTAGT	TATCTTATAT	TCTGTTGAG	CGTGTGAG
45	132119	AT	TTTGCCCTTTC	TCATCTTTTT	TAGTAATGCT	TAATTCTGCT	TTGTGATCAA	CTTTTTTACT	ATTTGTCT
	132189	TC	GTGATACCCC	CGACAGAATA	TTTTTTCAAT	TGATATTTAT	TGTCCTCTAA	AACGATAAAT	ACATCAAT
	132259	AT	TATCGTATGG	CCCATCTTTA	TATTTCTTAT	CATCTTTTCC	AACTAAAGCT	ATTTTATAAA	TGAACCAA
	132329	TC	TGGTATAACA	TTCAATAACC	TAACGTGCTG	CCATGGTTTC	AGCATAAAGC	CAAATTGCTT	TTCAGAAT
	132399	CT	AAACTCGGTT	TCGTATAATA	TGTTCTTAAA	TCCTTATATT	TAGGGTTTAT	TTCTTTTGTG	GCTTGTGT
50	132469	TG	TGGTTGGCGA	TTGCGGTTGT	TTTGTTCGTA	CATTAGATGA	ATGCGGTGTT	GTTGGGTTTG	GTGTTGGT
	132539	TG	TGGCGATTGC	GGTTGTACAT	TTGGTGAAGG	CGGAGTTGTT	ACACTTGGTG	TTGGTTTGGG	CGATTGCG
	132609	GT	TGTACATTGG	GTGAAGGCGA	TGTTGTTACT	TTTTGTATTT	TAGATTTTAT	TGCCTGTTCT	GTTTTAAG
	132679	TG	CTGATGCTAC	TTCGTCATTA	GCCGATTGCG	TTGTTACTAT	ACTTGCACCT	GTTGTTAAAA	GTGATAGT
	132749	GC	TAAACTGGTT	TTGCGCAATT	GTGTCATTTT	CATAGTTGTA	TGCTCCAATC	TATTATATTC	GATTGTTT
55	132819	TT	TACGTAATTA	GAATCATACA	ACTACATTAT	AGAAGACGTA	TCGGTCTATT	CACATTAAAC	CATGTTTA
	132889	AC	TATATTTGAA	TCATCGTTAA	GTAATCCATA	ATTTTCGTTA	AATACAGATG	GTTTCGAAGC	GATTCAAC
	132959	GC	TTGCACATTG	GGCATTGCAT	CATTTTCAAG	ACATTGAGAT	GTCACCTTTA	CAAACAAGTG	AATAATTT
	133029	CC	CTTAACATCA	TTACGCAAAA	CAAAAGAAGC	TAAGCAATGA	TGTAGGTCAAT	TGTCGCTTAA	CTTCTCAT
60	133099	TT	CATCGTTCAA	CGGTTCTATG	TAGTATCCAT	CACGATGATT	TTAGATTCAC	TTCAATTCTA	TCAATATT
	133169	GT	TACCATCAAT	AACATCTGCC	ATGCGATGTT	GTTGCAATTT	TTTTGTGTAAT	TCAAACGTGT	ACTTTCCA
	133239	CC	ATTTTTCATT	TTAATAAGTA	TTGTTCTCTG	ACCAATATTA	CCGTACAAAT	TATGTTGTTT	AATGAGTT
	133309	GT	TTTCTCAATT	TAAATCAAG	TTCTTTCAAG	GAAATCTCTT	CTTTAGTAAAT	CTTTGATTCT	GAAACATC
	133379	AT	GTGAGATTGT	ACCTTTTTTA	TCTTCTTTAG	TAATACTTAT	TCCAGTTTGT	TAATCAACTT	TCTTTCTA
	133449	TT	TGCTTTTGTG	ATACCACCGA	CCGAGTATTT	GTCTACGCCA	TATTTATTTT	GTTCTAAAAA	GATAAATA
65	133519	CA	TCGACATGCC	TATGCACGCC	TTCAACATAT	TTCTTATCAT	CTTTTCCAAC	TAAAGCTATT	TTATATAT
	133589	GA	AATAATCTGG	AACAATATTC	ATAAATCTTA	TTGTTGTCCA	TTTTTTGATA	ATTATACCCA	ACTCATTT
	133659	TT	AAATTTCTAA	CTTGATTTCC	TATAATATGA	TCTTAAACCT	TTAAATTTAG	GGTTTATTTT	CTTTTGTG
	133729	CT	TGTTTTGTGG	TTGGCGATGT	TGGTTGTGGT	GTTTCGACAT	TAGATGAAGG	CGGTGTTGTT	ACACTTGG
	133799	TG	TTGGTTGTGT	TGATTGCGGT	TGTGCATTTG	GTAAGACTG	AGTTGTTCCG	CTTGTGTGCT	GTGTGTGT
70	133869	GC	TTTTTGTATG	TTAGATTTTG	GCGTCTGTPC	TGTTTGAGTT	GAAAGCGATG	TTGATACATT	TTCTGTGT
	133939	TA	GACTTTGCGG	CCTGTTTCAG	ATTAGCCGTA	TGATTTTGTG	CTGCGTTTCA	TGCTTGCCTG	GTCGTTGT
	134009	GA	TTGCGCTTGT	TGATACGTTT	ATCTTTGCTA	ACTGTCTCTG	TTTAAGTGCT	GATGCTACTT	CTGCATTG
	134079	GC	CGATTGCGTT	GTTACTATAC	TTGCACCTGT	TGTTACAAGC	CCTAACGCTA	TACTCGCTTT	CGCAATTG
	134149	TT	CTTATTTTCA	TAGTTGTATG	CTCCAATCGT	AATTATTCGA	TTGTCTTTTA	CGTAATTAGA	ATCATACA

134219 AC TTCATTATAG AGGACATATT GGTCTATTCA CATTAAACCCT CGTTTAACAA TTGTTGAAAT ATTATTAA  
 134289 GT CATTTTAAGCA AAAAATAATG AGTACTTTCG AGGTTTATGT GGTGTTACTA TGTATCTTTG AAATGATT  
 134359 TT AAGACATAAA AAAGAAGTTA AGTAACGTTT TATCATCACT TAACCTCACT ATTGCAATTT CAATTCAT  
 134429 TA CGACGACCAG TCTAAGCATT ATGCTTTTTT AACTTTGACT TCAATCCTAT CAATATCTAT GACTTTGA  
 134499 CG ATATCGGCCA TTCGGTCTTC TTGTAACCTT TTATCCAATT CAATTGTATA CTTGCCATAG TATTTTCAT  
 134569 TT GGACAGTAAC TGTTCCTGAA CTCATTTTAC CATAAAGACC ATAATGATCA ATAAGATGTT TTCTTATT  
 134639 TT AAAATCAACC TCTTTTAAATG TCATTGCTTC TTTATCTATT TTAAATGGGA AAAAGTCATA ATCATATT  
 134709 CA CCAGTGTGAT CTTCTTTAAT GACTCTTGCT TCTGCTACTT GGTGACAAAC TTTATCGCTT GCTTTTGT  
 134779 GA TACCACCAAT CGAGTACTTT GCACCTTCAA ATCTCTTATC CTCATTAACA TAAAATATAT CAAGGTTG  
 134849 CG ATGTACACCC GTATGATAGT GTTCCTTATC TTTGCCAACT AAAGCTAGAT TATTAACCTG ATTACCTT  
 134919 CT ATCATATTCA AAAATTTAGT GCCTGGTTTA GGTGGAATAC TACTGATATA ACCTGTCACA TTTTATTA  
 134989 TT CAACACTAGG TTGAGTATAG TACGCTTTTA AATTTTACTT ATTTTGGCTT ACTTGAACAT GTTTATTT  
 135059 TC GGCATGCACC GGTGTGTTCCG TTACAGTATT TACACCTGTA GCTAAAATAC CCAACACTAA ACTTACTT  
 135129 TT GCAATTGATT TCAATTTTCAT AATTGTATGC TCCAATATAT TAGATTAGAT TGTTTTATTA CGTAAATT  
 135199 GA ATCATACAAC CATATTATAG GAGCTGTATG CTGATATTCA CATTAAACCCT TTTTAACTA TTCATAAA  
 135269 AT ATGATTAAGC TAGTTAAGTA AAAGATAATG CGTGCTGCGA CTGTTTGTGT GAGTTTGTGA ACTTTGCC  
 135339 CT CAAAAAGAA GCTAAGCAAC AGTACGTCGC TTAACCTCCA TATATTTCCA TAACAAGCAC GATTCAAC  
 135409 AT AGCGCTTTGT CGTTAGAAAG TATTATTCA TTTCTACTAG AATTTTGTGT ATTTGTCTGC CGTCGATA  
 135479 GA ATCACCATA CGTTCCTTTT CAAGTTTGTG ACTTAAATCG ATAGTATGTG ATTTGCCATC TTTTCATAG  
 135549 TA ATTGTAATTT GACCTTGTGT AAGACCATTG GAATACAATC CGTGTGTGTT GATTGCACGT TCTCGTAA  
 135619 TC TATAATCAAG TTCTTTTAAAT GAGATATTTT CTTTATAAAT TTGATACAGA CTATTTTGTG GTTCTTCA  
 135689 TT ACCTTTCTTA ATTTCAAGGA TTGGAGAGTG TATATAGTCG TAATACACTC CTCTATTAGT TTTTGTTA  
 135759 CG CCACCAATCG AATGATTATC AGCTTGTCTA CCGTTTCTCT CTCTAACGAC AAACACATCT ATATCTGG  
 135829 GT TAACTCTTTC ATTATATTTA TCTTTATCAG AACCAACTAA AGATATTTTA TTATATTGTC CATTGTAT  
 135899 AT AAAATCAATG AATCTATCTT TACCTTCAAT AAATCCATAT CCAGTTACTT TTTTATATTC AAAACTTG  
 135969 GT CCATTATAAT AAAGTTTAA ATCAGCCTCA CTTTGTGTTT TAACTTCTGT TTTTCGCTGT ACTGATTG  
 136039 TA TATTCGACGT TATTACACCT GTTGCTAACA TTCCCAATGC TAAACTTGCT TTTGCTATCG CTTTAAAT  
 136109 TT CATAATTTTT AGCCCCAATC TATTTAAATT TGATTTTTC AATACGTAAT ATGAACAATC TACCTACA  
 136179 TT ATATTGAGGC GCTATTCCCT TTCAAATTA ACCTCATTTA ACTAACTTG AATCATTGTT AAGGTAAT  
 136249 TA ATAAATGCT AACTGTAAAT AATCAATTCT TATCCGTCAA GACTCTTTT TATACAATTC ACACTTTA  
 136319 AA CCATTTTAAAT TTTAAAACT GTAAACCTT CAGACACAAA AATGACATCA TATCGTCGCA GTACCATG  
 136389 TC AAAATTTTCA AAGATCATGG TGACACGCAT CACTAAAAC GCGTCATTAC GGTATTTTAA ACATTAAT  
 136459 AA CATATGCGTA TTTACTATAC ACGGCGACAA TAAATACAGA TTGCAATTTT TTATAAAATC AAATGTTT  
 136529 TG AAAATCATCG AATAAGTATT ATGTTATGTA TATAAAGATT AGGAGATGAA GTGAATGAAT ATTATGTT  
 136599 AA CAGGTGCTAC AGGTCATTTA GGCACACATA TTCAAATCA AGCCATTGCA

## SEQ ID No:54 - SSL1

40 gene 135433..136128  
 /gene="EMRSA 16(252) ssl1"  
CDS complement 135433..136128  
 /gene="EMRSA 16(252) ssl1"  
 /product="(similar to SET6 from strain Mu50) SSL1"  
 45 /translation="MGLKIMKFKAIKASLALGMLATGVITSNIQSVQAKTEVKQQSEADLKLYYNGPSFEYKQVT  
 GYGFIIEGKDRFIDFIYNGQYNKISLVGSDKDKYNEEVNPDIDVFFVREGNGRQADNHSIGGVTKTNRGVYYDYIHS  
 PILEIKKGNEEPQNSLYQIYKENISLKYELDYRLRERAIKQHGGLYSNGLKQGGQITITMKDGKSHTIDLSQKLEKERM  
 50 GDSIDGRQIQKILVEMK"

## SEQ ID No:55 - SSL2

55 gene 134449..135150  
 /gene="EMRSA 16(252) ssl2"  
CDS complement 134449..135150  
 /gene="EMRSA 16(252) ssl2"  
 /product="(similar to SET7 from strain N315) SSL2"  
 60 /translation="MKMKSIKVSLLVLGILATGVNTVTEQPVHAENKHVQVSQNSKNLKAYYTQPSVEYKQVTGYI  
 SSIQPKPGTKFLNMIEGNTVNNLALVGKDKYHYHTGVHRNLDFVYVNEKRFEGAKYSIGGITKASDKVVDQVAAE  
 RVIKEDHTGEYDYDFFPKIDKEAMTFLKEVDKIRKHLIDHYGLYGEMSSGTVTVQMKYQKYTIELDKKLQEDRM  
 65 ADIVKVIDIDRIEVKVKKA"

## SEQ ID No:56 - SSL3

gene 133130..134161  
 /gene="EMRSA 16(252) ssl3"

CDS complement 133130..134161  
 /gene="EMRSA 16(252)ss13"  
 /product="(similar to SET9 from strain N315)SSL3"  
 5 /translation="MKIRTIKASIALGLVTTGASIVTTQSANA EVASALKAGQLAKINVSTSAITTTAQAVNAEQ  
 NHTANPEQAAKSNTENVSTSLSTQTEQTPKSNIQKATQPAPSGTTQSLPNAQPQSTQPTPSVTPPSSNVETPQPT  
 SPTTKQAQKEINPKFKGLRSYYTKSSLEFKNELGIIKKWTTIRFMNIVPDYFIYKIALVGKDDKKYGEVHRHVD  
 VFIVLEQNKYGVGDKYSVGGITKANRKKVDYKTGISITKEDKKGTISHDVSEYKITKEEISLKELDKFLRKQLIEQH  
 10 NLYGNIGSGTIVIKMKNKGKTYTFELHKKLQHRMADVIDGTNIDRIEVNLKSS"  
  
SEQ ID No:57 - SSL4  
gene 131863..132783  
 /gene="EMRSA 16(252)ss14"  
 15 CDS complement 131863..132783  
 /gene="EMRSA 16(252)ss14"  
 /product="(similar to SET8 from strain Mu50)SSL4"  
  
 /translation="MKMTQIAKTSALSLTGTGASIVTTQSANA EVASALKTEQAISKIQKVTTSPPSPNVQPQSP  
 20 QPTPSVTPPSPNVQPQSPQPTPNPTTPHSSNVETKQPQSPTTKQAQKEINPKYKDLRTYVTKPSLEFEKQFGFML  
 KPWTTVRFMNVIPDWFIYKIALVGKDDKKYKDGPDNDVFIIVLEDNKYQLKKYSVGGITKTNSKKVDHKAELSIT  
 KKDEKGI SHDDSEYKITKEEISLKELDKFLRKQLIEQHNLYGNIGSGTIVIKTKNGGKYTFELHKKLQHRMADV  
 25 IDGTSIERIE  
 VNLKSS"  
  
SEQ ID No:58 - SSL5  
gene 130798..131502  
 /gene="EMRSA 16(252)ss15"  
 30 CDS complement 130798..131502  
 /gene="EMRSA 16(252)ss15"  
 /product="(similar to SET3)SSL5"  
  
 /translation="MKMTAIAKASLALSILATGVITSTAQTVNASEHESKYENVTKDIFDLRDYYSRASKELKNVT  
 35 GYRYSKGGKHYLIFDKNRKFTRIQLFGKDIERIKRKNPGLDIFVVKAEENRNGTVYSYGGVTKKNQ GAYYDYL  
 PRFVIKKEVGAGVSVHV KRYIYKEEISLKELDKFLRQYLIQDFDLYKKFPKASKIKVTMKDGGYYTFELNKKLQ  
 NRMSDVIDGRNIEKIEANIR"  
 40 SEQ ID No:59 - SSL7  
gene 129656..130351  
 /gene="EMRSA 16(252)ss17"  
 45 CDS complement 129656..130351  
 /gene="EMRSA 16(252)ss17"  
 /product="(similar to SET1)SSL7"  
 /translation="MKLKTAKATLALGLLTGTGVTSEGQAVQAAEQERVQHLHDIRDLDHRYSSSESFEYSNVSG  
 KVENYNGSNVVRFNPKDQNHQLFLGKDKQYKEGLQGQNVFVQELIDPNGRLSTVGGVTKKNNKTSETNTPLFV  
 50 NKVNGEDLDASIDSFLIQKEEISLKELDKFKIRQQLVNNYGLYKGT SKYKIIINLKDENKVEIDLGDKLQFERMGD  
 VLNSKDIRGISVTINQI"  
  
SEQ ID No:60 - SSL9  
gene 128617..129315  
 /gene="EMRSA 16(252)ss19"  
 55 CDS complement 128617..129315  
 /gene="EMRSA 16(252)ss19"  
 /product="(similar to SET5)SSL9"  
 /translation="MKLTAIAKATLALGILTTGVM TAESQTVNAKVKLDETQRKYYINMLKDYYSQESYESTNISV  
 60 KSEDYYGNSVNLNFNQRNKNFKVFLIGDDRNKYKELTHGRDVFVPELIDTKGGIYSVGGITKKNVRSVFGYVSHPG  
 LQVKKVDPKDGF SIKELFFIQKEEVS LKELDFKIRKMLVEKYRLYKGASDKGRIVINMKDEKKHEIDLSEKLSFDR  
 MFDVLD SKQIKNIEVNLN"  
 65 SEQ ID No:61 - SSL10

gene 127571..128254  
/gene="EMRSA 16(252) ss110"  
CDS complement 127571..128254  
/gene="EMRSA 16(252) ss110"  
/product="(similar to SET4) SSL10"  
/translation="MKLTALAKVTLALGILTTGTLTTEAHSGHAKQNKSVNKHDKALHRYTGNFKEMKNIN  
ALRHGKNNLRFKYRGMKTQVLLPGDEYRKYQORRHTGLDVFFVQERRDKHDISYTVGGVTKNKTSGFVSTPRNLN  
TKEKGEDAFVKGYPDYDIKKEEISLKEIDFKLRKHLIEKYGLYKTLKDGRIKISLKDGSFYNLDLRTKLKFKHMG  
VIDSKQIKDIEVNLK"

SEQ ID No:62 - SSL11

gene 123447..124145  
/gene="EMRSA 16(252) ss111"  
CDS complement 123447..124145  
/gene="EMRSA 16(252) ss111"  
/product="(similar to SET15 from strain Mu50) SSL11"  
/translation="MKLKNIKASLALGILTTGMITTTAQPVKASEQSRLSVTSNDTQELKKYSGTGYNFQNV  
SGYREKDKMNIIDGTQLNVVTLTGTDKERFKDYDYEGLDVFFVREGSGKQAEINISIGGITKTNKNNDYKDFVNNV  
GLEITKPTGHNTATRQAEYRINKEEISLKEIDFKLRKHLIENHLYKTEPKDGKIRITMKGGGYTFELNKLQ  
HRMGDVIDGRNIEKIEVDLY"

SEQ ID No:63

38500 A CAATGTAAGT TGGGGTGGGG CCCCAACATA GAGAAATTGG GTTACCAATT TCTACAAGCA ATGCAAGTT  
38570 G GGGCGGGCCC AAACCCAGAG AATGACTTGG CATTTCTATT AGACTCCAAA TCGCATTTAA TATTCTGTG  
38640 T ACTAAATTAA TGATTGTGTA AGTACTATTG AGACATTTTG TAGTATTACT CAAATATTGT TTATCCCTTA  
38710 G TATTTACGAT ATAAATCATA CGACCTTATT TAATCAATGA GGACTTCGAC TGTTATTTTT TGAGGATTT  
38780 T TAACATAAAC ATTTGCAAGA TCAGAATGTA ATCTTTTACT CAAATCAATT GTAAAGTCAT TGCCGCCAC  
38850 C GGTGATTTTA AGTTTACCTT TATTGCGATT TTATGATAT AATATTTTAT TTTTAATTAG TGTTTCACG  
38920 T GCTCGAAAAA CAATTTCTTT CAATGTTAAT ATTGGTTTAT TACCTTGGTA CATCTTATGT GCACCAATA  
38990 A TGTGTTGTAG TTTATCTTTG TATTGGACAA AAAGACTATA TGGTTTATCA GAAGGTGATG CAGCAGGTG  
39060 T TACACCACCT GTAAATGTAT CACGATATAC CCACCATAAC TGGCTAGTGT CTTCGTCTTT TAAACCAAA  
39130 C ACGTCTACAT AGCGGTCTCT TAATTGATTA ACGTTGCCCC AACTTTCAGC ACCCCATAAT CGTATATGC  
39200 G CTGACCAAGA ATAACGCTTA AATTCAACAT AAATGTTGCC GTCATCATAT TGGTATAACC ATCTATTTG  
39270 A AAATGTGAGA GGCAGTTGTG TGAATATTTT AATTAACCTA TTGATGTTCC TTTCTGCTTG GCCAATATG  
39340 A TAAGCTTTAG CGTCTGTAGA AACATTTAAT GGCGAATTAG AGGTTTGTGT AGCTGTTGTT CCTAATAAT  
39410 A ACAATGTGTG CATTAATAAT AACTTTTTCA TAATGCACCT TCTCATTCGA ACTTCTCCTT TCTGCCGTAA  
39480 T CATCAATACG AGAATTAACA AATTTAAAAA TGAACGGCGT CAATGTAAAC ATTGTGCTTA TTTCTGTTT  
39550 A CTTCGATTTT GATGTTAGCT AGATTTTTTC AGGAGGATTT CGATGTGTGC AGTTTGAGGG TTTTAAACA  
39620 T ATCTATTTGC ATCAGTTGAT TTTAATCTTT TGCTTAAATC TATTGTGTAG TTTGTGCTGC CACCTGTAA  
39690 T CTTAATGTGG CCTTTATTAT AATCACCATT ATATAATTTT TTGTTTTTAA TTAGTGTGTT GCGGATTTCG  
39760 G AAATCAACTT CCTTTAAAGT TAAAACTGGT TTATTACCTG TATAAAATTG ATATTCGCCA ATAAATGTT  
39830 C CTGCTTTATC TTTATATGTT ACATTCAATT TATAATGTTT ATCGCTAGGT TTTGCTACTG GTGTAACAC  
39900 C GCCAGTAAAT GTTCTTGGG CTAACATATA AGTGCGCCAG GTTTCATAAT CTTTTAAAGC AAAAATATC  
39970 T ACATATTTAT TTCTTAATTG ATTAATATTA CCCCAACTTT CAGGCGCGAA TACTTGAATA TGACTAAC  
40040 C ATGATAACTG TAACAATGTT GCATGGATG TACCATTATC TTTTGTCCAT AGCCATTTAT TTGAAATG  
40110 T TAAATGGCGT TGTGTGTAAT ATTTAGTCAA TTCATTTACA TTCGTTTCGC TTTCACTTAT ATTGTATGC  
40180 T TTTGCTTCAG ATGACATACT AATCGGTGAA TGAGGGGATT GACTTGTAGC TGTGCTATC ATAGTCAAT  
40250 G CTATTGATAA AATAATTTTT TTTGTGAGGT TGTTTTTCAT TAGAAAATCT CCTTTACGAG GAATGTTAT  
40320 A TGCGAGGAAT AATAAGTTTA TAAAAAGAGA GAAATCAATT AAATCATTCA GTCATTTTGT ATTCACCTA  
40390 A TGAATGATAC TAATGAGATT TTCAAGAAAT GACTTCAATT TCTGCATGTC GAGGATTTTT AACATAACT  
40460 G TTTGTGTGAG TTAGTTTTAA CTTTTTACTC AAATCAATCG TATAATGATT ACCACCACCT GTAATCCTA  
40530 A TTTGACCTTT ATTAAATCTT CCATTATATA GCTTTTTGTT CTTAATTAAA GTTTGACGTA CGCGGAAAT  
40600 C TAGTCTTTTT AAAGTTATTA CTGGTTTTAT ACCTTGATAA AACTCAAGTC CACCAATCAT AGTTTGTG  
40670 T TTATCCTTAT ATTTTAAAAA TAATTTATAA GGCTTATCAG ATGAAGTAGC TGCTGGTGTT ACACCTCCA  
40740 G TAAAGGTTTC ATCATAAGTC CAATAGCCTT CTATAGTATC CTCATCCTTG GTACCAAAAA TATCTACAT  
40810 A TTTATCTCTT AATTGATTAA TGTTGCCCCA ACTTTCAGGA CCAAATACTT GAATATGCGT GTACCAAAC  
40880 C CACGTTTGTG ATGTTGCATG TATGGTACCG TTAGGGTTTT GCCATAGCCA CTTTCCAGAA AATGAAAGG  
40950 T GTGGCTGCGT GTAATATTTG ATTAATTCAT TGATATTCGT TTCGTTTTCA CTTATATTGT ATGCTTTTG  
41020 C TTCAGATGAA AAAGTATTTG GTGCTTTTAG CGGTGTGTGT GATGTAGTTC CTAAAGTAA CAATGCTGT  
41090 T GATAAAACAA TTTTGTTCGT GATGTTCTTT TCCATATGAT AATCTCCTTT GCGTGAATTA CCCATAGTA  
41160 T ATAAGGTATT GCACCGATTG GGAATTAAT AAAGCTAAC ACTATGTTAA ATAAACATAA ACAGTTGAA  
41230 C GTGTGATTTA AGTAAATTTT TTCATTTTTA AGCATTTGGA CATTGCCAA TATAAAATGA AAAGAGTGC  
41300 T AATAAATTAA AACTGATGCT AATTATGATA TTAATTGCGT TAAATTAGTT ATTTATTGAA TTTTCTAAC  
41370 T AACTGAAAAA TAGTTATACT TTAATGTAG

SEQ ID No:64 - SSL12

gene 40400..41125

*S. aureus* strain MSSA-476 taken from unpublished genome project at the Sanger centre via <http://pedant.gsf.de>

31

	112210	A	AGCATAACAC	ACATACTTGA	CGACGAAATA	ATTTGAAATT	GAAATAGAGA	GGTTAAGTGA	CGATCAAAC
	112280	G	TTGCTTAACT	TCCTTTTAAAT	GCTTAAAAAT	CATTTCAAAG	GCACATAGAA	ACGCTATATT	AACCTCATA
	112350	A	TCACTCATT	TTTTTTGCTT	AAATTACTTA	ATAATACTTC	AATAATTGTT	AAAAGGGGTT	TAATGTGAT
5	112420	T	ATCTTAGAAT	GCCATCTATA	ATGATGTTGT	ATGATTCAAA	TTACGTAAAA	AGACAATCGA	ATATAATAT
	112490	A	GATTGGAGCA	TACAATTATG	AAAATGAGAA	CAATTACTAA	AACCAGTTTA	GCACAGGGC	TTTTAACAA
	112560	C	AGGCGCAATT	ACAGTAACGA	CGCAATCGGT	CAAAGCAGAA	AAAAACAAAT	CAACTAAAGT	TGACAAAGT
	112630	A	CCAACGCTTA	AAGCAGAGCG	ATTAGCAATG	ATAAACATAA	CAGCAGGTGC	AAATTTCAGC	ACAACACAA
	112700	G	CAGCTAACAC	AAGACAAGAA	CGCACGCCTA	AACTCGAAAA	GGCACCAAAAT	ACTAATGAGG	AAAAAACCT
10	112770	C	AGCTTCCAAA	ATAGAAAAAA	TATCACAACC	TAAACAAGAA	GAGCAGAAAA	CGCTTAATAT	ATCAGCAAC
	112840	G	CCAGCGCCTA	AACAAGAACAA	ATCACAACAG	ACAACCGAAT	CTACACGCAC	GCAAACTAAA	ATGACAACA
	112910	C	CTCCATCAAC	AAACACGCCA	CAACCAATGC	AATCTACTAA	ATCAGACACA	CCACAATCTC	CAACCATAA
	112980	A	ACAAGCACAA	ACAGATATGA	CTCCTAAATA	TGAAGATTTA	AGAGCGTATT	ACACGAAACC	GAGTTTGA
	113050	A	TTTGAAAAAGC	AGTTTGGATT	TTTGCTCAAA	CCATGGACGA	CGGTTAGGTT	TATGAATGTT	ATTCCAAAT
	113120	A	GGTTCACTA	TAAAATAGCT	TTAGTTGGAA	AAGATGAGAA	AAAAATATAA	GATGGACCTT	ACGATAATA
15	113190	T	CGATGTATTT	ATCGTTTTAG	ATGACAATTT	ATATCAATTG	AAAAAATATT	CTGTCCGTGG	CATCACGAA
	113260	G	ACTAATAGTA	AAAAAGTTAA	TCACAAAGTA	GAATTAAGCA	TTACTAAAAA	AGATAATCAA	GGTATGATT
	113330	T	CACGCGATGT	TTCAGAAATAC	ATGATTACTTA	AGGAAGAGAT	TTCCCTGAAA	GAGCTTGATT	TTAAATTGA
	113400	G	AAAACAACCTT	ATTGAAAAAC	ATAATCTTTA	CGGTAACATG	GGTTTCAGGAA	CAATCGTTAT	TAAAAATGA
20	113470	A	AACGGTGGGA	AATATACGTT	TGAATTACAC	AAAAAAGTCG	AAGAGCATCG	TATGGCGATC	TACGATGAT
	113540	G	GCACTAATAT	TGATAACATT	GAAGTGAATA	TAAATAATC	ATGACATTCT	CTAAATAGAA	GCTGTCATC
	113610	G	GAAAAACAAG	AAGTTAAGTG	ACAACGGTTT	ACATGTTGCT	TAGCTTCTTT	TATTATGCGT	AATGATGTA
	113680	A	AAAGACGAAT	ATTCAATTTGT	TTGTAAGAGT	GGCATTTCTA	TGCTTAAAAA	GTGACGAAAC	TTCAAAATG
	113750	G	CCAAGTGTG	AATCACATCA	AAATCATTTT	TATTTAACGA	ACATTATGGA	TTTCTTAATT	TACTTAACG
25	113820	A	TGATTCAAAT	ATAGTTAAAC	AAGGTTTTAAT	GTGAATGGAG	CAATGCGCCA	TCTATAATAA	AGCTGTATG
	113890	A	TTCAATGAAT	GTAATCGAAC	AAATCTAATA	ATTACGAATG	GAGCATACAA	CTATGAAAAA	AACAACAAT
	113960	T	GCTAAAACAA	GTTTAGCACT	AGGCTTTTTA	ACAACAGGTG	TAATCAACAAC	GACAACGCAA	GACGCAAAAT
	114030	G	CGACAACACC	GCCTTCAACT	AAAGTGGAAA	CACCGCAACA	AGTAGCAAAAT	GCAACAACAC	CATCTTCAA
	114100	C	TAAAGTGGAA	GCACCGCAAC	AAGCAGCAAA	CGCGACAACA	CCATCTTCAA	CTAAAGTAGA	AGCACCGCA
30	114170	A	TCAAAAACCAA	ACGCGACAAAC	ACCATCTTCA	ACTAAAGTGG	AAGCACCGCA	ACAAGCAGCA	AACCGGACA
	114240	A	CACCACCTTC	GTCTAATGTA	GACACATCAC	CACCACAATC	GCCAACCACA	AAACAAGTAC	CAACAGAAA
	114310	T	AAATCCATAA	TTTAAAGATT	TAAGAGCGTA	TTATACGAAA	CCAAGTTTAG	AAATTTAAAA	TGAGATTGG
	114380	T	ATTATTTTAA	AAAAATGGAC	GACAATAAGA	TTTATGAATG	TTGTCCCAGA	TTATTTTATA	TATAAAAT
	114450	G	CTTTAGTTGG	TAAAGATGAT	AAAAAATATG	GTGAAGGAGT	ACATAGGAAT	GTCCGATGAT	TTGTCTGTT
35	114520	T	AGAAGAAAAT	AATTACAATC	TCGAAAAATA	TTCTGTCGGT	GGTATCACAA	AGAGTAATAG	TAAAAAGT
	114590	T	GATCACAAGG	CAGGAGTAAG	AATTACTAAG	GAAGATAATA	AAGGTACAAT	CTCTCATGAT	GTTTCAGAA
	114660	T	TCAAGATTAC	TAAAGAACAG	ATTTCTTTAG	AAAGAACTGA	TTTTAAATTG	AGAAAACAAC	TTATTGAAA
	114730	A	AAATAATCTG	TACGGTAACG	TTGGTTCAGG	TAAAAATTGT	ATTAAAAATG	AAAACGGTGG	TAAGTACAC
	114800	G	TTTGAATTGC	ACAAAAAATT	ACAAGAAAT	CGCATGGCAG	ATGTCATAGA	TGGCACTAAT	ATTGATAAC
40	114870	A	TTGAAGTGAA	TATAAAATAA	TCATGACATT	CTCTAAATAG	AAGCTGTCAT	CGGAAAAACA	AGAAGTTAA
	114940	G	TGACAACGGC	CTACATGTTG	CCTAGCTTCT	TTTGTATGTT	TCGATGATTG	TAGAACCCGA	ATTTTCGAT
	115010	G	GGTCCAAATA	TGACGTGGAA	GAGTCCTGAA	TTTATCTGTA	AATCCCTATC	TATCGGGTGT	GGAGCACAA
	115080	C	GGATCAGTTT	TTATTTAACG	AACATTATAG	ATTCCTTAAT	TTACTTAATA	ATGATTCAAT	GATTATTAA
45	115150	A	CATGGTTTAA	TGTGAAAGGT	CAAATACGCC	AAATATAATA	AAGCTGTATG	ATTCAATAGA	CGTAAGCGA
	115220	A	CAAACTAAT	AATTACGAAT	GGAGCATAAC	ACTATGAAAA	TGGCAGCAAT	TGGCAAGGCA	AGTTTAGCA
	115290	T	TAGGTATTTT	AGCAACAGGA	ACAATAACGT	CATTGCATCA	AACGTGTAAT	GCGAGTGAAC	ATGAAGCAA
	115360	A	ATATGAAAAA	GTGACAAAAA	ATATCTTTGA	CTTAAGAGAT	TACTATAGTG	GCGCAAGTAA	GGAACTTAA
	115430	A	AATGTTACTG	GTTATCGTTA	TAGCAAAAGT	GTTCAAGCATT	ACCTTATCTT	TGATAAACAT	CAAGAGTTC
50	115500	A	CTAGAATACA	AATTTTTGGT	AAAGATATAG	AAAGATTTAA	AGCACGCAAA	AATCCGGGAT	TAGACATAT
	115570	T	TGTTGTTAAA	GAAGCGGAAA	ACCGTAATGG	CACAGTGTTC	TCATATGGTG	GTGTCACATA	GAAAAATCA
	115640	A	CAGCGTTATT	ATGATTATAT	AAACGCACCA	AGATTTCAAA	TCAAGAGAGA	TGAAGGTGAC	GGTATTGCT
	115710	A	CGTACGGTAG	AGTACACTAC	ATTTATAAAG	AAGAGATTTC	ACTTAAAGAA	CTCGACTTTA	CTCAGGACAC
	115780	A	GTATTTAATT	CAAAATTTTG	ATCTGTATAA	AAAGTTTCCT	AAAGATAGTA	AGATAAAAGT	GATAATGAA
	115850	A	GATGGCGGCT	ATTATACGTT	TGAACCTAAT	AAAAAATTAC	AAACAAATCG	CATGAGTGAC	GTCAATTGAC
55	115920	G	GTAGAAATAT	TGAAAAAATA	GAAGCCAACA	TTAGATAATG	CAATGAAATA	TGGATAATAG	TAAAAATG
	115990	G	ATAGTATAGA	GGAGTTAGGC	AACATAAGTT	GCTTAGCTTC	TTTTTTGTGT	TGGAGAGATG	AAAAAGAAG
	116060	C	GTATCGATGA	ATAATAAAAA	CACCAATAAA	ACTTGTGGGA	ATAGTTGATA	CCTATAGTCG	CGCGTTGTC
	116130	C	TTTTCGTGAC	ATGAAACAAT	GTGGAAAAAC	TAATTAATTT	GAGGGAAAGT	GTGAATAGTT	AAAAAAGCT
60	116200	G	CGTTAAGTTT	AAAAAATAGA	TTAACGCTGT	TAGGATTCTA	TTAATTAGCT	TAACATTGGT	TCAAAAATA
	116270	G	TTAAAAAGAG	GTTAATTCAT	AGCTTAGTAT	TCCGCTTATA	TAATGATAGT	AGATTGTTTC	TATTACGTA
	116340	A	TTGAAATAAT	CATATAAAAA	TATATTAAAG	CAAAATTTAT	AAATAGATTG	GGAGAATAGT	ACTATGAAA
	116410	T	TAAAGCGGTT	AGCTAAAGCA	ACATTAGTAT	TGGGATTGTT	AGCTACTGGT	GTAAATAACA	CAGAAAGTC
	116480	A	AACAGTAAAA	GCGGCAGAAAT	CAACTCAAGG	TCAACACAAT	TATAAATCAT	TAAAAACTAC	CTATAGCAA
	116550	G	CCAAGTATAG	AGTTAATAAA	TGTAGATGGT	CTGTATAGAC	AACATTTAAC	TGATAAAGGT	GCATATGTA
65	116620	T	GGAAAAATCT	TAAAGATTAT	TATATTGGGC	TACTAGGTGA	AGATAGTAAG	AAATTCAAAT	CAGATGTAT
	116690	A	CGGGGACCTA	GATGCATTTT	TAGTCTATAGA	AGAAGAACCT	GTTAAAGGAA	GACAAATATC	AATTGGCGG
	116760	T	ATAAGTAAGA	CAAATAGTAA	AGAATTTAAA	GAAAGAGAA	TCGATGTTAA	AGTAACAAGA	AAAGCAGAC
	116830	A	GAGATACTAC	ATCAACTAAA	GATAGTAAAT	TTAAAAATTAC	AAAAGAAGAA	ATCTCGTTAA	AAGAGTTAG
	116900	A	TTTTAAATTA	AGACAAAAAT	TGATGAAAGA	AGAGAAATTTA	TACGATGCAA	TTAACCATAG	AAAAGGTAA
70	116970	A	ATTTGATGTTA	AAATGGAAGA	TGATAAGTTT	TACTATTTCG	AACTTACAAA	AAAATACAAA	CCGATCGC
	117040	A	TGGGTGACAC	GATAGATGGT	ACCAAAATCA	AAGAAATTA	TGTTGAGCTA	GAATATAAAT	AATCTTTGG
	117110	A	CAAGCAGACT	AGTAATTGTA	GGGAAGTTAA	GCGGATAACAT	ATTGCTTAGC	TTCTTTTTTG	TTTTTTTAT
	117180	G	ATGAAAAAAG	GAGCGGGTTT	ATGATCAAGT	TTTTGGAAAA	ACGGTTGATA	CTTATAGTCG	CGCGTTGTC
	117250	C	TTTTCGTGAC	ATGAAACAAT	GTGGAAAAACA	TAATTAATTT	GAGGGAAAGT	GTGAATAGTT	AAAAAATTA

117320	G	TATTGTGTTA	TAAAAAATAA	TTAATACTGT	TAGGATTTC	TTAACTA	TAACGTTGGT	TCAAAAATA
117390	G	TTAAAAAGAG	GTTAATTCAT	AGCGCAGTAT	CTCAGTTTATA	TAATGATAGT	AGATTGTTCCG	TATTACGTA
117460	A	TTGAATTAAT	CATATAAAAA	TATATTAAGA	CAAAATTTAT	AAATAGATTG	GGAGAATAGT	ACTGTGAAA
117530	T	TAAAAACGTT	AGCTAAAGCA	ACATTGGCAT	TAGGCTTATT	AACTACTGGT	GTGATTACAT	CAGAGGCC
117600	A	AGCAGTGCAA	GCAAAAGAAA	AGCAAGAGAG	AGTACAACAT	TTATATGATA	TTAAAGACTT	ACATCGATA
117670	C	TACTCATCAG	AAAGTTTTGA	ATTCAGTAAT	ATTAGTGGTA	AGGTTGAAAA	TTATAACGGT	TCTAACGTT
117740	G	TACGCTTTAA	CCAAGAAAAT	CAAAATCACC	AATTATTCTT	ATCAGGAAAA	GATAAAGATA	AATATAAAG
117810	A	AGGCCTTGAA	GGCCAGAAATG	TCTTTGTGGT	AAAAGAAATTA	ATTGATCCAA	ACGGTAGACT	ATCTACTGT
117880	T	GGTGGTGTA	CGAAGAAAAA	TAACCAATCT	TCTGAACTA	ATACACCTTT	ATTTATAAAA	AAAGTGTAT
117950	G	GCGGAAATTT	AGATGCATCA	ATTGAATCAT	TTTTAATTAA	TAAAGAAGAA	GTTCACCTGA	AAGAAGCTTG
118020	A	TTTCAAAAT	AGACAACATT	TAGTTAAAAA	TTATGGTTTA	TATAAAGGTA	CGACTAAATA	CGGTAAGAT
118090	C	ACTTTCAATT	TGAAAGATGG	AGAAAAGCAA	GAAATTGATT	TAGGTGATAA	ATTGCAATTC	GAGCACATG
118160	G	GCGATGTGTT	GAATAGTAAG	GATATTCAAA	ATATAGCAGT	GACTATTAA	CAAATTTAA	GTAAGTAAT
118230	C	AATGACTCTA	AAGTAATAAA	TTTGAAGCAG	CTTAGCGATG	AAATGTTGAA	TAGATACGTA	CACCTTACA
118300	T	AAAGGAGCGT	ATTTAAAAACA	ACCTTGTCGT	TAGGCTTTTT	TTACGTTT	TAACGCGAGG	TATGAGCGT
118370	A	CTAAAAATTC	ACATTACTTC	TGAAAGTGAT	GTCCATTGAA	TATTAATTAG	TTCTTCATTA	ACCATGATT
118440	T	AAATTTAATT	AAACGAGTTC	TAAATGTCAGT	CTGTCTCAAT	GCCCTTTATA	ATAAATGTGT	ATTATTCAA
118510	A	TTACGTAATA	AAAGCAATCC	AATATATTAA	GATTGGAGCA	TATGAATATG	AAATTTACAG	CGATAGCTA
118580	A	AGCGATATTT	GTATTAGGAA	TATTAACAAC	AAGTGTAAATG	ATAACAGAAA	ATCAATCCGT	TAATGCAAA
118650	A	GGAAAGTATG	AAAAAATGAA	CCGTTTATAT	GATACAAACA	AGTTACATCA	ATACTATTCA	GGACCTAGT
118720	T	ATGAGTTAAC	AAATGTTAGT	GGCCAAAGTC	AAGGTTTATTA	TGACTCTAAC	GTTTTGCTTT	TTAACCAAC
118790	A	AAATCAAAAG	TTCCAAGTGT	TTTTATTGGG	AAAAGATGAA	AATAAATACA	AAGAAAAAAC	ACATGGTTT
118860	A	GATGTCTTTG	CGGTACCGGA	ATTAGTAGAT	TTAGATGGAA	GAAATTTTAG	TGTTAGTGGT	GTAAACAAAG
118930	A	AAATGTGAAA	ATCAATATTT	GAGTCTCTAA	GAACGCCGAA	CTTACTAGTT	AAAAAATAG	ACGATAAAG
119000	A	CGGTTTTTCG	TATGATGAAT	TTTTCTTTAT	TCAAAAGGAA	GAAGTATCAT	TGAAGGAACT	TGATTTCAA
119070	A	ATAAGAAAAC	TGTTAATTTA	AAAAATACAA	TTGTATGAAG	GGGCAGCTGA	TAAAGGTAGA	ATTGTTATT
119140	A	ATATGAAAGA	TGAAAATAAG	TATGAAATG	ATTTAAGTGA	TAAATTAGGT	TTGAGCGTA	TGSCAGATG
119210	T	CATTAATAGT	GAACAAATTA	AAAACATCGA	AGTGAATTTG	AAATAATCAA	TGATATATAT	AGAAATGAAA
119280	G	CTTAAGAAGC	GGTTAATAAA	TCCCATGTTT	AATGATTTTG	ATACGTGTTT	TAATAATAAA	AACATATCG
119350	A	ACATTGACTA	CGTTATTAAAG	CTGCTTTTTT	GTACACTTTG	TATCGAATAA	CTTAAGATCT	AAAACATAA
119420	C	GGAAGAAGAA	ATGATTCCTCC	TAAAAAAAT	TATGTTGCTA	TTAAAAATCA	GTTAATACGA	ATGTTAACA
119490	T	ACGTTTGATT	TTCATTAAATA	ATGATTCAAG	TTTATTTAAA	TGAGCGTTAA	TGTCAGTCTG	TTTTGATGC
119560	A	CCTTATAATA	AAGACAGATA	GTTCAAATTA	CGTAATAATA	ACAATCCAAT	ATATCAAGAT	TGGAGCAAA
119630	T	AAATATGAAA	TTTACGGCAT	TAGCAAAAGC	AACATTAGCA	TTAGGAATAT	TAACTACAGG	TGTTGTTAC
119700	A	ACAGAAAGTA	AAGCTGTCTCA	CGCGAAAGTA	GAACCTGATG	AGACACAACG	CAAATATTAT	ATCAATATG
119770	C	TACATCAATA	CTATTCTGAA	GAAAGTTTTG	AACCAACAAA	TATTAGTGTT	AAAAGCGAAG	ATTACTATG
119840	G	CTCTAACGTT	TTAAACTTTA	AACAACGAA	TAAAGCTTTT	AAAGTATTTT	TACTTGGTGA	CGATAAAAA
119910	T	AAATATAAAG	AAAAAACACA	TGGCCTTGAT	GTCTTTGCG	TACCTGAATT	AATAGATATA	AAAGGTGGC
119980	A	TATATAGCGT	TGGCGGTATA	ACAAAGAAAA	ATGTGAGATC	AGTGTTTGGA	TTGTAAAGTA	ATCCAAGTC
120050	T	ACAAGTTAAA	AAAATCGATC	CTAAACATGG	CTTTTCGATA	AATGAGTTGT	TCTTTATTTA	AAAGGAAGA
120120	A	GTATCGTTGA	AGGAACTGGA	TTTTAAATA	AGAAAAATGT	TAGTCGAAAA	ATATAGATTG	TATAAGGCC
120190	G	CGTCAGATAA	AGGTAGAATC	GTTATTAAAT	TGAAAGACGA	AAAGAAATAT	GTAATTGATT	TAAAGTAAA
120260	A	ATTAAGTTTT	GATCGTATGT	TTGATGTAAT	GGATAGTAAG	CAAAATTAATA	ATATTGAAGT	GAATTTGAA
120330	T	TAAATTAAGTA	TAATAACTTA	AGAAGCGACT	TAACGACAAA	ATGTGAATTG	ACATGCATGT	CCTTAAATA
120400	A	GGAACTGTGT	TAAATACATT	ACTGTTGCTA	AGTTGTTTTT	TGCGTTTCAA	AGAGCAGAAC	AGAGTAACA
120470	T	CATCAGTTGT	AGTAAACGAT	AATCTAGTAA	AACAACATAA	TGAAATAATG	AAATTCATTT	AACCTGAAC
120540	A	TTAAATATATA	TTTGTTTTTTC	ATTAAGAATA	ATTCAGATAT	ATTTAAATCG	AGGTTAATTA	TCGTATGAA
120610	A	CGATGCACGT	TATAATAAAA	ATGTATGATT	CAAAATACGT	AATGAAACAA	ATCCAATATA	TTAAGATTG
120680	G	AGCAAAACAA	TATGAAATTA	ACAGCGATAG	CTAAAGCTGC	ATTAGCTTTA	GGAAATTTAA	CAACAGGAA
120750	C	TTTAACAACA	GAAGTTTCATT	CAGGTCATGC	AAAACAAAAT	CAAAAGTCAG	TAAATAAACA	TGACAAGGA
120820	A	GCATTATACC	GATACTACAC	TGAAAGAGCT	ATGGAAATGA	AAAATATTAG	TGCTTTGAAA	CATGGTAAA
120890	A	ATAACTTGCG	TTTTAAGTTT	AGAGGTATTA	AGATTCAAGT	TTTACTGCCT	GGAAATGATA	AAAGTAAAT
120960	T	TCAACAGCGT	AGTTATGAGG	GGTTAGATGT	TTTTTTTGTT	CAAGAAAAAA	GAGATAAGCA	CGATATATT
121030	T	TATACTGTTG	GTGGTGTAAT	ACAGAATAAT	AAAACATCTG	GAGTTGTCAG	TGCACCAATA	TTAAATATT
121100	T	CAAAAGAAAA	GGGTGAAGAT	GCTTTTGTTA	AGGGTTACCC	TTATTACATT	AAAAAGAAA	AAATAACAC
121170	T	AAAAGAGCTG	GATTATAAGT	TGAGAAAGCA	TCTAATCGAA	AAATATGGAC	TTTATAAAAC	AATCTCAA
121240	A	GATGGTAGGG	TCAAAATTTAG	CTTGAAAGAT	GGCAGTTTTT	ATAACCTTGA	TTTAAAGTCT	AAATTAATA
121310	T	TCAAATATAT	GGGGGAAGTC	ATAGAAAGCA	AACAAATTA	AGATATTGAA	GTAAACTTAA	AGTAAATAA
121380	T	TACGAATAAT	AAAAAGTAAT	TGAAGCGGCT	TAAACGATGA	AAGTAAATTG	ATGCGCATAC	CTTACCAA
121450	A	GGATGCATCA	ATCGATATCG	TCGTTAAGCT	TTTTTGTTT	ACGTTTCATG	GATTTCTACC	CAATTTTCA
121520	T	AAATATAAAA	ATTCCACCAC	CAACATCAAA	ATTCTCAACA	TCGCAACATA	CCCAATGTT	ATAATAAAT
121590	C	TATTACACAA	AGAGATAAAT	TACTTATTTA	AAGGCGGAGG	AATCACATGT	CTATTACTGA	AAAACACAG
121660	T	CAGCAACAAG	CTGAATTACA	TAAAAAATTA	TGGTCGATTG	CGAATGATTT	AAGAGGGAAC	ATGGATGCG
121730	A	GTGAATTCGG	TAATTACATT	TTAGGCTTGA	TTTTCTATCG	CTTCTTATCC	GAAAAAGAC	AACAAGAAAT
121800	A	TGCAGATGCG	TTGGCAGGTG	AAGATATCAC	GTATCAAGAG	ACATGGGCAG	ATGAAGAATA	TCGTGAAGA
121870	C	TTAAAGCTG	AATTAATTGA	TCAAGTCGGT	TACTTCATTG	AACCACAAGA	TTTATTCAGT	GCGATGATT
121940	C	TGGAATTTGA	AACGCAAGAT	TTGCAATATC	AACATCTGGC	GACGGCAATT	CGTAAAGTTG	AACAATCAA
122010	C	ATTAGGTGAA	GAAAGTGA	ATGACTTTAT	CGGACTGTTC	AGCGATATGG	ACTTAAAGTTC	AACGCGACT
122080	A	GGTAACAATG	TCAAAGAAGC	TACTGCTTTA	ATCTCTAAAG	TCATGGTTAA	TCTTGACGAT	TTACCATTT
122150	G	TTCCAGTGTA	TATGGAAATT	GATATGTTAG	GTGATGCATA	TGAATTCCTA	ATTGGGCGCT	TTGCGGCGA
122220	C	AGCAGGTAAA	AAAGCAGGCG	AGTTCTATAC	ACCACAACAA	GTATCTAAGA	TACTGGCGAA	TACTGTCTAC
122290	A	GACGGTAAAG	ATAAATTACG	TCATGTGTAC	GACCAACAT	GTGGTTCCGG	TTCATTATTG	TTACGTGTT
122360	G	GTAAGAAAC	GCAAGTGTAT	CGTTATTTCG	GTCAAGAACG	TAACAATACC	ACTTACAAC	TAGCAGCGA

122430 T GAACATGTTA TTACATGATG TACGTTATGA AAATTTTCGAT ATCCGTAATG ATGACACGTT GGAAATCC  
 122500 A GCCTTTTATG GACATACATT TGATGCGGTT ATTGCGAACC CACCATACAG TCGGAAATGG ACAGCAGAT  
 122570 T CAAAATTTGA AAATGACGAA CGATTACAGCG GATACGGCAA ACTTGCGCCA AAGTCCAAAG CAGACTTTG  
 5 122640 C CTTTATTCAA CACATGGTAC ATTACTTAGA CGATGAAGGT ACCATGGCCG TTGTACTCCC ACATGGTGT  
 122710 C TTATTCCGTG GTGCTGCAGA AGGTGTCTATT CGTCGTTATT TAATTGAAGA AAAGAAGTAC TTAGAAGCC  
 122780 G TGATTGGCTT ACCAGCCAAT ATTTTCTATG GGACAAGTAT TCCAACATGT ATTTTAGTAT TTAATAAAT  
 122850 G TCGCCAACAA GAAGACTATG TATTATTTAT CGATGCATCC AATGATTTTG AAAAAGGAAA AAATCAAAA  
 122920 C CATTAAACCG ATGCCCAAGT CGAACGCATT ATTAACACAT ATAAGCGTAA GGAACAATT GATAAATAT  
 122990 A GCTACAGTGC GACATTACAA GAGATCGCCG ATAACGATTA CAACTTAAAC ATACCGAGAT ATGTCGATA  
 10 123060 C ATTCGAAGAA GAAGCACCAG TGTGATTTAGA TCAAGTCCAA CAAGATTTGA AAAATATCGA CAAAGAAAT  
 123130 C GCAGAAGTTG AACAAGAAAT CAATGCATAC CTGAAAGAAC TTGGGGTGTG GAAAGATGAG TAATACACA  
 123200 A AAGAAAAATG TGCCAGAGTT GAGGTTCCCA GGGTTTGAAG GCGAATGGGA AGAGAAGAAG TTAGGGGAC  
 123270 C TTAACACAA AATAGGTAGT GGAAGACTC CCAAGGTGG AAGTGAAGAC TATACAAACA AAGGCATAC  
 123340 C ATTTTAAAGG AGTCAAATA TTAGAAATGG TAAATTAAAT CTTAATGACT TAGTTTATAT TAGTAAAGA  
 15 123410 T ATAGATGATG AGATGAAAAA TAGTAGAACG TACTATGGTG ATGTTCTTTT AAATATTACA GGAGCATCA  
 123480 A TAGGTAGAAC AGCCATTAAAT TCGATAGTTG AAATACATGC TAATTTAAAT CAACATGTAT GTATTATTA  
 123550 G ATTGAAAAAA GAGTATTATT ATAATTTTTC TGACAGTAT CTATTATCAA GAAAGGTAA AAGGAAAAT  
 123620 T TTCCTTGCAC AAAGTGGAGG TAGTCGAGAA GGACTAACT TCAAAGAAAT TGCTAATTTA AAAATCTTC  
 123690 A CCCCACATAT ATTTGAAGAG CAGCAAAAAA TAGGCGAATT CATCAGCAAA CTTGACCGAC AAATTTGAAT  
 20 123760 T AGAAGAACAA AAACCTGAAT TACTTCAGCA ACAGAAAAAA GGCTATATGC AGAAATCTT CTCGCAAGA  
 123830 A TTGCGATTCA AAGATGAGGA AGGTAAAGAT TATCCAGATT GGAATCATA ATCAATTCAA GAAATATTT  
 123900 G AGAATAAGGG TGGCACTGCT CTAGAAGACG TAATTTAATT TGACGGTAAT TATAAGTTA TAAGTATAG  
 123970 G AAGTTATTCT ATAAATAGCA CTTATAATGA TCAAAATATA AGAGTCAATA AAAATAAAA AACTGAAAA  
 25 124040 A TATATTTTAT CAAAAGGCGA CTTAGCAATG GTATTAAATG ATAAACAAA AGATGGGAAA ATTATAGGT  
 124110 A GAAGTATATT TATAGATAAA GATAATCAAT ATATTTATAA TCAAAGAACT GAAAGATTAA TACCATTTG  
 124180 C TGAAAATGAT AATAAATTTT TATGGTTCTT AATGAATACA GATTTAATTA GAAATAAAT AAAAGGTAT  
 124250 G ATGCAAGGAG CAACCCAAGT TTATATAAAT TATTCATCTA TTAAATTGAT ATCTATACAA TTGCCACTT  
 124320 C TTGAAGAACA ACAGAAAAATA AGAGGGTTTC TAGAAGTTT ATCTGGAATA ACTACTAAC AATTGCACA  
 30 124390 A GATAGACCAA TTAAGAGAGA GGAAAAAGGC GTTTTACAG AAAATGTTT TTTGATTTGT CGCTGCAAT  
 124460 A TAGTTTTTAT TATTTGTTTA TTTAGATGCT TACCCATCA TATGCGTAA CTTTACAAA TAAGAAATA  
 124530 A AGTTCAATGA AATCAAAAC GAACATTAAA TTTAGGCACT GTGATAGCAC AGTGTCTTTT TTGTGTCGA  
 124600 A ATTGCTGACA GAATAGTAG TTAATAAAG ATTAAGTTGA GATAAAGTGT TATTCGTAA TAAAGAGA  
 35 124670 G TAGATCGATA GGAATTGAAT GATATTAGT AACTATTAT TAAATTACTT AATAATTACT AATTTTAG  
 124740 T TAAAGTAAGT TTAATGTGAA GCACGACCAT TGCTCATTAT AATGAATGAG GATTGTTCTG ATTGCGTAA  
 124810 T AGAATAAATC AAATAGACTA AAAATTGGGA GCATAGAATT ATGAAATTAA AAAATATTGC TAAAGCAAG  
 124880 T TTAGCACTAG GGATTTTAA ACACAGGATG ATTACAATA CTGCTCAGCC AGTAAAGCAA ATTGAGCAA  
 124950 A GCAGATTATC AGTTACTTCA AAAGATTACAC AAGAATTAAA AAAATACTAC AGTGGAAACAG GATATAATT  
 40 125020 T TCAAAATGTG AGTGGTTATA GAGAAGGTAA TAAATGAAC ATTATTGATG GACCACAAC TAATGTAGT  
 125090 T ACTTTACTTG GCACAGACAA AGAAAGGTTT AAGGACGATG AAGATTATGA AGGACTTGAT GTATTTGTT  
 125160 G TAAGAGAAGG GTCAGGTAAA CACGCAGATA ATATATCAAT TGSTGGAATT ACAAACAAA ATAAGAATC  
 125230 A ATATAAAGAC CCTGTACAAA ACGTTAATTT ATTGACTTCT AAGAGTAACG GTCAAAATAC TGCTTCTGT  
 125300 G ACTTCAGAAT ACTATAGCAT CAATAAAGAA GAAATTTTCA TAAAGAACT TGATTTCAA CTAGAAAG  
 125370 C AATTAAATGA TAAACATGAT CTTTATAAGA CAGAGCCTAA AGACAGCAA ATTAAGTTT CTATGAAA  
 45 125440 A TGGCGGCTAC TATACGTTTG AATTAAATA AAAATTACAG CCTCATCGCA TGGGTGATAC GATTGATAG  
 125510 T AGAAATATAA AGAAATTTGA AGTGAATTTA TAATAATATT CGAGGGAGTA TATCATGAGA GAAATTTT

SEQ ID No: 68 - SSL1

50 gene 110541..111236  
 /gene="MSSA-476ss11"  
CDS 110541..111236  
 /gene="MSSA-476ss11"  
 /product="(Similar to SET6 from strain Mu50)SSL1"  
 /translation="MGLKIMKFKAIKASLALGMLATGVITSNVQSVQAKTEVKQQSEADLKLYNGPSFEYKKVT  
 55 GYGFIEGKDRFIDFIYNGQYNKISLVGSDKDKYNEVNPIDIVFVREGNGRQADNHSIGGITKTNRGVYDYIHT  
 PILEIKKGKEEPQSSLYQIYKEDISLKELDFKLKQLISQSGLYSNGLKQGQITITMNDGTHHTIDLSQKLEKERM  
 GESIDGRQIQKILVEMK"

60 SEQ ID No: 69 - SSL2

65 gene 111522..112217  
 /gene="MSSA-476ss12"  
CDS 111522..112217  
 /gene="MSSA-476ss12"  
 /product="(Similar to SET7 from strain N315)SSL2"  
 /translation="MKMKSIIVKISLLLGILATGVNTTTEKPVHAEKKPIVISENSKKLKAYYTQPSIEYKNVTGYI  
 SFIQPSIKFMNIIDGNSVNNIALIGKDKQHYHTGVHRNLNIFYVNEDKRFEGAKYSIGGITSANDKAVDLIAEARV

IKADHIGEYDYDFFPFKIDKEAMSLKEIDFKLRKYLIDNYGLYEMSTGKITVKKKYYGKYTFELDKKLQEDRMSD  
VINVTIDIDRIEIKVRKA"

5 SEQ ID No:70 - SSL3

gene 112508..113578  
/gene="MSSA-476ss3"

10 CDS 112508..113578  
/gene="MSSA-476ss13"  
/product="(Similar to SET8 from strain Mu50)SSL3"

15 /translation="MKMRTITKTSALGLLTGAVTTTQSVKAEKIQSTKVDKVP TLKAERLAMINITAGANSAT  
TQAANTRQERTPKLEKAPNTNEEKTSASKIEKISQPKQEEQKTLNISATPPAPKQEQSQTTESTTQOTKMTTPPST  
NTPQPMQSTKSDTPQSPTIKQAQDTMPKYEDLRAYYTKPSFEFEKQFGFLLKPWTTVRFMNVIPNRFIYKIALVG  
KDEKKYKDGPDYDNIDVFIVLEDNKYQLKKYSVGGITKTNSKKVNHKVELSTKKDNQGMISRDVSEYMITKEEISL  
KELDFKLRKQLIEKHNLGNMGSGTIVIKMKNNGGKYTFELHKKLQEHMADVIDGTNIDNIEVNIK"

20 SEQ ID No:71 - SSL4

gene 113943..114890  
/gene="MSSA-476ss14"

25 CDS 113943..114890  
/gene="MSSA-476ss14"  
/product="(Similar to SET9 from strain N315)SSL4"

30 /translation="MKITTIKTSALGLLTGAVTTTQAANATTPPSTKVETPQQVANATTPSSTKVEAPQQA  
NATTPSSTKVEAPQSKPNATTPSSTKVEAPQQAANATTPSSNVDTSPPQSPTTKQVPTEINPKFKDLRAYYTKPS  
LEFKNEIGIILKKWTTIRFMNVVPDYFIYKIALVGKDDKKYGEVHRNVDVFVLEENNYNLEKYSVGGITKSNSK  
KVDHKAGVRITKEDNKGITISHDVSEFKITKEQISLKELDKLRKQLIEKNNLYGNVSGKIVIKMKNNGGKYTFELH  
KKLQENRMADVIDGTNIDNIEVNIK"

SEQ ID No:72 - SSL5

35 gene 115254..115958  
/gene="MSSA-476ss15"

CDS 115254..115958  
/gene="MSSA-476ss15"  
/product="(Similar to SET10 from strain Mu50)SSL5"

40 /translation="MKMAAIAKASLALGILATGTITSLHQT VNAS EHEAKYENVTKDIFDLRDY  
YSGASKELKNVTGYRYSKGGKHLYLFDKHQKFTRI QIFGKDIERFKARKNPGLDIFVVKAEENRNGTVFSYGGVTK  
KNQDAYYDIYINAPRFQIKRDEGDGIATYGRVHYIYKEEISLKELDKLRQYLIQNFDLYKKFPKDSKIKVIMKDGG  
YYTFELNKKLQTNRMSDVIDGRNIEKIEANIR"

45 SEQ ID No:73 - SSL6

gene 116404..117102  
/gene="MSSA-476ss16"

50 CDS 116404..117102  
/gene="MSSA-476ss16"  
/product="(Similar to SET2 from strain NCTC6571)SSL6"

55 /translation="MKLKALAKATLVGLLATGVITTESQT VKAESTQGQHNYKSLKYYYSKPSIELINVDGLYR  
QHLTDKGAYVWKNLKDYYIGLLGEDSKKFKSDVYGD LFAFLVIEEPVKGRQYSIGGISKTNSKEFKEREVDVKVT  
RKADRDTTSTKDSKFKITKEEISLKELDKLRQKLMKEENLYDAINHRKGKIVVKMEDDKFYTFELTKKLQPHRMG  
DTIDGTIKIEINVELEYK"

60 SEQ ID No:74 - SSL7

gene 117524..118219  
/gene="MSSA-476ss17"

65 CDS 117524..118219  
/gene="MSSA-476ss17"  
/product="(Similar to SET1-C)SSL7"

/translation="MKLKTAKATLALGLLTTGVITSEGQAVQAKEKQERVQHLYDIKDLHRYSSSEFEFSNISG  
KVENYNGSNVVRFNQENQNHQLFLSGKDKDKYKEGLEQNVFVVKELIDPNGRLSTVGGVTCKNNQSSETNTPLFI  
KKVYGGNLDASIESFLINKEEVSLKELDFKIRQHLVKNYGLYKGTTKYGKITFNLKDGEKQEIDLGDKLQFERMGD  
VLNSKDIQNIQIAVTINQI"

SEQ ID No:75 - SSL8

10        gene                    118558..119256  
                                 /gene=" MSA-476ss18"  
                                 /gene="MSSA-476ss18"  
                                 /product="(Similar to SET12 from strain Mu50) SSL8"  
15        /translation="MKFTAIAKAIFVLGILTTSMITENQSVNAKGKYEKMNRLYDTNKLHQYYSGPSYELTNVSG  
                                 QSQGYDSNVLLFNQONQKFQVFLGKDKENYKEKTHGLDVFAVPELVLDGRIFSVSGVTCKNVKSIFESLRTPN  
                                 LLVKKIDDKDGFYSYDEFFFIQKEEVSLKELDFKIRKLLIKKYKLYEGAADKGRIVINMKDENKYEIDLSDKLGFER  
                                 MADVINSEQIKNIEVNLK"

SEQ ID No:76 - SSL9

25        gene                    119635..120333  
                                 /gene=" MSA-476ss19"  
                                 /gene="MSSA-476ss19"  
                                 /product="(Similar to SET13 from strain Mu50) SSL9"  
30        /translation="MKFTALAKATLALGILTTGVFTTESKAVHAKVELDETQRKYINMLHQYYSEESFEPTNISV  
                                 KSEDYYSNVLFNFKQNKAFKVFLLGDDKNYKEKTHGLDVFAVPELVLDGRIFSVSGVTCKNVRSVFGFVSNP  
                                 LQVKKIDPKHGFSINELFFFIQKEEVSLKELDFKIRKMLVEKYRLYKGASDKGRIVINMKDEKKYVIDLSEKLSFDR  
                                 MFDVMSKQIKNIEVNLN"

SEQ ID No:77 - SSL10

35        gene                    120692..121375  
                                 /gene=" MSA-476ss110"  
                                 /gene="MSSA-476ss110"  
                                 /product="(Similar to SET14 from strain Mu50) SSL10"  
40        /translation="MKLTAIAKAALALGILTTGTLTTEVHSGHAKQNKSVNKHDKALRYRYTGKTMEMKNISAL  
                                 KHGKNNLRFKFRGIKIQLVLLPGNDKSKFQQRSYEGLDVFFVQEKRDKHDIFYTVGGVIQNNKTSVGSAPILNISK  
                                 EKGEDAFVKGYPYIKKEKITLKELDYKLRKHLIEKYGLYKTISKDGRVKISLKDGSFYNLDLRSKLKFKYMGEVI  
                                 ESKQIKDIEVNLK"

SEQ ID No:78 - SSL11

50        gene                    124851..125543  
                                 /gene=" MSA-476ss111"  
                                 /gene="MSSA-476ss111"  
                                 /product="(Similar to SET15 from strain Mu50) SSL11"  
55        /translation="MKLKNIASLALGILTTGMITTTAQPVKAIEQSRLSVTSKDTQELKKYSGTGYNFQNVSG  
                                 YREGNKMNIIDGPQLNVVTLGTDKERFKDDEYEGLDVFFVREGSGKHADNISIGGITKTNKNQYKDPVQNVNLL  
                                 TSKSNGQNTASVTSEYYSINKKEISLKELDKFLRKQLIDKHDLYKTEPKDSKIKVSMKNGGYTTFELNKKLQPHRM  
                                 GDTIDSRNIKKIEVNL"

SEQ ID No:79

60        62000 C CTAAATTGTA AGCGCATACA AAATAAACAC AACCTACTAT TAAAATTGTT AATATTTTAT CAATAATTA  
65        62070 A ATGAACATTT TATTAATATT AAATTTAAGT AGTAGGAAAT AATTAAAATA AGTACTACAT TTAAAGTAT  
                                 62140 A ACTATTTTTC AAGTACTTAG AAAATTCAT TATCAACAA TTTAATGCAA TTGATTAGAG AATAATTGT  
                                 62210 A ACGTGTCGTT TTTAATATAT AACTCCGCC TACTTTATTA AGTACTGTTT CTGTCCAAA CTTAAAAA

5  
10  
15  
20  
25  
30  
35  
40  
45  
50

```

62280 G ATAAGTTTGT CTTAAATAAC ACTACTAACT GTTTAAGTTT ATTTAACATA GTTTTAGCTT TTATTTAAT
62350 T CCGAATCGGT GTAATAGCTT ATATACTTTG GGTAATTCAC GCAAAGGAGA TTTTCATATG AAAAAGAAC
62420 A TCATGAATAA ATTAGTTTTA TCAACAGCAT TGTTACTTTT AGGAACATA TCAACACAAC TTCCTAAAA
62490 C ACCAATCAGT TTTTCATCTG AAGCAAAGC CTATAATATC AGTGAAAACG AGACTAATAT CAATGAGTT
62560 A ATAAAGTATT ATACACAGCC TCATTTATCA CTATCAATAA AATGGTTATG GCAAAAGCCC AATGGTAGC
62630 A TTCATGCAAC ATTGCAAACG TGGGTTTGGT ATAGTCATAT TCAAGTGTTC GGATCCGAGA GTTGGGGAA
62700 A CATTAAATCAG TTAAGAAATA AATACGTTGA TATATTTGGA ACTAAAGATG AGGACACAGT TGAAGGTTA
62770 C TGGACTTATG ATGAAACATT TACTGGTGGT GTTACGCCAG CAGCTACTTC ATCTGATAAG CCTTATAGA
62840 C TATTTTAAAA ATATAGTGAT AAACAACAAA CTATCATCGG TGGACATGAA TTTTACAAAG GAAATAAAC
62910 C AGTATTAAC TTAAGAAGAT TAGATTTCCG TATTCGTCAA ACATTAATAA AAAATAAAAA GTTATATAA
62980 C GGAGAATTTA ATAAAGGTCA AATTAAGATA ACTGCTGATG GAAATAATTA CACGATTGAT TTAAGTAAA
63050 A AGTTAAATTT AACTGACACA AACCGTTATG TTAATAATCC TAAAAATGCA CAAATTGAAG TCATACTCG
63120 A AAAATCTAAC TaACCTATTA CCTTTTGTAA ATGCGGATaA TTTCAAttat TTTCAAttat CTTTATATA
63190 A TTAAACATTC CaacaatTACT CAAAGGAGaa AttCGAAga acAATaacaT CaCGaAAAAA ATTATTTTA
63260 T CaaCAACATT GTTACTATTA GGTACAGCAT CTACACAATT TCCTAATACA CCTATCAATT CTTCTACTG
63330 A AGCGAAAGCT TATTATATAA ATCAAAACGA AACTAACGTT AATGAGTTAA CTAATATTTA CTCGCAAAA
63400 A TATTTAACCT TCTCTAACAG TACGTTATGG CAAAAGATA ACGGTACGAT TCATGCAACG TTGTTACAG
63470 T TTTCTTGGTA TAGTCATATT CAAGTTTATG GACCTGAAAG TTTGGGCAAT ATCAACCAAT TAAGAAATA
63540 A AAGCGTTGAT ATTTTGGGCA TAAAGACCA AGAAACCAT GATTCTTTTG CATTATCTCA AGAACCGTT
63610 T ACTGGTGGTG TTACTCTGTC AGCAACATCT AACGATAAAC ACTATAAACT GAATGTGACA TATAAGAT
63680 A AAGCAGAAAC GTTTACTGGC GGATTTCCAG TTTATGAAG CAATAAGCCT GTTTTAACTT TAAAGAAT
63750 T AGATTTTTCGT ATTCGTCAAA CATTAATTAA AAGTAAAAA TTATATAATA ATCTTTATAA TAAAGGACA
63820 A ATTAATAATA CAGGTGCAGA CAATAAGATT ACAATAGATT TAAGTAAAG GTTGCCATCA ACTGATGCA
63890 A ATAGATATGT TAAAAACCT CAAAATGCAA AAATTGAAGT TATCCTCGAA AAATCAAACT AACATAAT
63960 A ATGGAGTTAA TAAAAATAAT CGCAAAACT ATATTGACTT CGCTCACATT TAAATTTCTT ATTCCTCGT
64030 A TCATGATTCC TCTGAAAGGA GATGTTCTAA TGAGTAAGAA CATCACGAAA AATATAATTT TAACGACAA
64100 C ATTATTACTA TTAGGTACTG TATTACCTCA AAATCAAAA CCAGTATTTA GTTTTACTC TGAAGCTAA
64170 A GCTTATAGCA TTGGTCAAGA TGAACCTAAC ATCAATGAAT TAATTAAATA TTACACACAG CCTCATTTT
64240 T CATTTTCAAA TAAATGGCTA TATCAATATG ATAATGGAAA CATTTATGTT GAACCTAAGA GATATTCAT
64310 G GTCAGCACAT ATATCTTTAT GGGGCGTGA AAGTTGGGGA AATATTAATC AGTTAAAAGG TCGTTACGT
64380 A GATGTGTTTG GACTAAAAGA CAAAGATACT GATCAGTTAT GGTGGTCTTA TAGAGAGACA TTTACAGGT
64450 G GCGTTACACC AGCCGCAAAA CCTTCTGATA AACTTATATA TCTTTTGTG CAATACAAAG ATAACTAC
64520 A AACGATTATT GGTGCGCATA AAATATACCA AGGCAATAAA CCAGTATTTA CATTGAAAGA AATCGATT
64590 C CGTGACAGAG AAGCGTTAAT AAAAAATAAA ATATTATATA ACGAAAATCG TAAATAAGGT AAGCTTAAG
64660 A TCACCGGTGG CGGTAATAAC TACACTATTG ATTTAAGCAA AAGATTACAT TCAGATCTAG CAAATGTTT
64730 A TGTAAAAAAT CCTAATAAAA TAAGTGTGA CGTCCTCTT GATTAGTATA TGAGGTGAC TTATCTTTC
64800 A TGCACCTTAA TTCCAAATCA GATTATTTAA ATGATAATTT TAAAGTGTA TGATGTATAT AATAGGTAA
64870 A ATTTTCTATA TATTTAAATG GAATTTGGG TAGGAATGTG ACAGAAATAG TATTTTATAA AATTTATTT
64940 C GTTGTCACAT CCCAATTGCT ATTGTCTGTA GAATTTCTT TTGAAATTTCT CTATGTTGGG GCCCGCCA
65010 A CTTGCACATT ATTGTAAGCT GACTTTTCTG CAGCTTCTGT GTTGGGGCCC CGCCTATAAT TGAATAATG
65080 C TTGTTACATG GGCATTTTCA TTCCGCTCAAC TACTACCAAT ATAATATTGT AGAGCCTAAG ACATTTGATT
65150 T ATTATGTCTT AGGCTCTATT CCTTCATTTA ATGATTCAAT TATTATAGCA ATACTTTATT GTCCCATGA
65220 T TAGTGTTCTT TTAATGAGAC ATAGTAACTA TAAAGTTTAA TAATCGTTCT AAATCTAGCA TCTTCTAGT
65290 T TGGTTTTCCC ATTTCTTAAA TCTTGTACTG TTTGATATGG AACTCCTGAA TTTTTCGAAA TTTTATAGC
65360 C TGATTAGAC TCAAACAAAG ATTGAAGAGA GTTTATAATA TTATTTCACT CTGTCTTTT TAATCCCT
65430 T TCTTGAATTA ACAATATATA ATGTTGTTAT TAAAACAGTC AGTGTATGGA TGATTTCAAT TCCTAAAA
65500 T

```

## SEQ ID No:80 - SSL12

55  
60

```

gene 62408..63133
      /gene="MSSA-476ss112"
CDS 62408..63133
      /gene="MSSA-476ss112"
      /product="(Similar to SAV1168 from strain Mu50)SSL12"

/translation="MKKNIMNKLVLSTALLLLGTTSTQLPKPTPIFSSEAKAYNISENETNINELIKYYTQPHLSL
SNKWLWQKPNNGSIHATLQTVWVYSHIQVFGSESWGNINQLRNKYVDIFGTDKEDTVEGYWYDETFPGGVTPAATS
SDKPYRLFLKYSQKQTIIGGHEFYKGNKPVLTLELDFRIRQTLIKNKKLYNGEFNKGQIKITADGNNTIDLSK
KLKLTDTNRYVKNPKNAQIEVILEKSN"

```

## SEQ ID No:81 - SSL13

65  
70

```

gene 63227..63952
      /gene="MSSA-476ss113"
CDS 63227..63952
      /gene="MSSA-476ss113"
      /product="(Similar to SA1010 from strain N315)SSL13"

```

5 /translation="MNNNITKKIILSTTLLLLGTASTQFPNTPINSSSEAKAYYINQNETNVNELTKYYSQKYLTF  
SNSTLWQKDNGTIHA TL LQFSWYSHIQVYGPE SWGNINQLRNKSVDFGIKDQETIDSFALSQETFTGGVTPAATS  
NDKH YKLNV TYKDKAETFTGGFPVYEGNKPVLT LKELDFRIRQTLIKSKKLYNNSYNKGQIKITGADNNYTIDLSK  
RLPSTDANRYVKKPQNAKIEVILEKSN"

10 SEQ ID No:82 - SSL14  
gene 64060..64776  
/gene="MSSA-476ssl14"  
CDS 64060..64776  
/gene="MSSA-476ssl14"  
/product="(Similar to SAV1166 from strain Mu50)SSL14"

15 /translation="MSKNITKNII LTTT LLLL GTVLPQNKPVFSFYSEAKAYSIGQDETINELIKYYTQPHFSF  
SNKWLYQYDNGNIYVELKRYSWSAHISLWGAESWGNINQLKGRYVDVFG LKDKD TDQLWWSYRETFTGGVTPAAKP  
SDKTYNLFVQYKDKLQTIIGAHKIYQGNKPVLT LKELDFRAREALIKNKILYNENRNKGK LKITGGGNNTIDLSK  
RLHSDLANVYVKNPNKITVDVLF D"

20 ***S. aureus* strain COL taken from unpublished genome  
project at The Institute for Genomic Research (TIGR) via  
ViroloGenome**

25 SEQ ID No:83

30 ATGAAATTTAAAGCGATAGCAAAAGCAAGTTTAGCATTGGGAATGTTAGCAACAGGTGTAATTACATCGAATGTACAATCAGTAC  
AAGCGAAAGCAGAAGTTAAACAACAAAGTGAATCAGAGTTAAAACACTATTATAATAAACCAATTTTAGAGCGTAAAAATGTGAC  
TGGATTTAAATATACTGATGAGGGTAAACACTATTTAGAAAGTCACAGTAGGGCAACAGCATTCTCGAATCACTTTACTTGGATCT  
GATAAAGATAAATTTAAAGACGGAGAAAACCTCAAATATAGATGTGTTATCCTTAGAGAAGGTGACAGTAGACAAGCAACAAATT  
ACTCAATTGGTGGCGTTACAAAATCAAATAGTGTGCAGTATATTGATTATATCAATACGCCAATTTTAGAAATCAAGAAAGATAA  
TGAAGATGTACTTAAAGATTTTACTACATTTCAAAGAAGACATCTCATTAAAAGAACTTGATTATAGATTAAGAGAACGTGCG  
ATTAAACAACACGGCTTGTATTCAAATGGTCTTAAACAAGGTCAAATTACAATTACAATGAATGATGGCACAACACATACAATCG  
ATTTAAGTCAAAAACCTTGAAAAGAAGCTATGGGTGAGTCAATCGACGGCCTAAGATTAATAAAATTTCTAGTAGAAATGAAATA  
A

35 SEQ ID No:84 - SSL1

40 gene 1..681  
/gene="COLssl1"  
CDS 1..681  
/gene="COLssl1"  
/product="(Similar to SET6 from strain N315)SSL1"

45 /translation="MKFKAIKASLALGMLATGVITSNVQSVQAKAEVKQOSESELKHYYNKPILERKNVTGFKYT  
DEGKHYLEVTVGQOHSRITLLGSDKDKFKDGENSNIDVFILREGDSRQATNYSIGGVTKSNSVQYIDYINTPILEI  
KKDNEDVLKDFYYISKEDISL KELD YRLRERAIKQHGLYSNGLKQGQITITMNDGTTHTIDLSQKLEKERMGESID  
GTKINKILVEMK"

50 SEQ ID No:85

55 ATGAAATTTAAAAATATTGCTAAAGCAAGTTTAGCACTAGGGATTTTAAACAACAGGGATGATTACAACACTACTGCTC  
AGCCAGTAAAAGCAAGTACATTAGAGGTTAGATCACAAGCTACTCAAGACTTGAGTGAATATTATAATAGACCGTT  
CTTTGAGTATACAAATCAGTCAGGATATAAAGAGGAAGGAAAAGTGACGTTTACTCCTAATTATCAACTTATAGAT  
GTAAC TTAACTGGGAATGAAAAGCAAAATTTTGGTGAAGATATTTCTAATGTAGATATATTTGTTGTAAGAGAAA  
ATTCGTAGATAGATCTGGTAATACAGCTTCAATTGGTGGTATTACTAAAACAAACGGTTCAAATTATATTGATAAAGT  
AAAAGATGTAAATTTAATAATTACTAAAACATCGATAGTGTTACATCAACGTCAACATCATCTACATATACAATT  
AATAAGAAGAAATTTTCAATAAAGAAGCTTGATTTTAAATTAAGAAAGCATTTAATTGATAAACATAACCTTTATA  
AGACAGAACCTAAGACAGTAAAATTCGAATTACTATGAAAGATGGTGGGTTCTACACATTTGAATTGAATAAAAA  
60 GTTACAAACACACCGTATGGGTGATGTTATTGATGGCAGAAATATAGAAAAAATTGAAGTGAATTTATAA

SEQ ID No:86 - SSL2

5 gene 1..678  
/gene="COLss12"  
CDS 1..678  
/gene="COLss12"  
/product="(Similar to SET7 from strain N315)SSL2"  
/translation="MKMKNIAKISLLLGILATGVNTTTEKPVHAEKKPIVISENSKKLKAYYNQPSIEYKNVTGYI  
10 SFIQPSIKFMNIIDGNSVNNIALIGKDKQHYHTGVHRNLNIFYVNEDKRFEKAKYSIGGITSANDKAVDLIAEARV  
IKEDHTGEYDYDFPFKIDKEAMSLKEIDFKLRKYLIDNYGLY GEMSTGKITVKKKYGYKYTFELDKKLQEDRMSD  
VINVTDIDRIEIKVIKA"

15 SEQ ID No:87

ATGAAAAATGAGAACAAATTGCTAAAACCAAGTTTAGCACTAGGGCTTTTAAACAACAGGCGCAATTACAGTAACGACGCAATCGGTCA  
AAGCAGAAAAAATACAATCAACTAAAGTTGACAAAGTACCAACGCTTAAAGCAGAGCGATTAGCAATGATAAACATAACAGCAGG  
20 TGCAAAATTCAGCGACAACACAAGCAGCTAACACAAGACAAGAACGCACGCCTAAACTCGAAAAGGCACCAAATACTAATGAGGAA  
AAAACCTCAGCTTCCAAAATAGAAAAATATCACAACTAAACAAGAAGAGCAGAAAAACGCTTAATATATCAGCAACGCCAGCGC  
CTAAACAAGAACAATCACAAACGACAACCGAATCCACAACGCCGAAACTAAAGTGACAACACCTCCATCAACAAAACGCCACA  
ACCAATGCAATCTACTAAATCAGACACACCACAATCTCCAACCATAAAACAAGCACAACAGATATGACTCCTAAATATGAAGAT  
TTAAGAGCGTATTATACAAAACCGAGTTTGAATTTGAAAAGCAGTTTGGATTATGCTCAAACCATGGACGACGGTTAGGTTTA  
25 TGAATGTTATTCCAAATAGGTTTCATCTATAAAATAGCTTTAGTTGGAAAAGATGAGAAAAATATAAAGATGGACCTTACGATAA  
TATCGATGTATTTATCGTTTGTAGAGACAATAAATATCAATTGAAAAAATATTCTGTCGGTGGCATCACGAAGACTAATAGTAAA  
AAAGTTAATCACAAAGTAGAATTAAGCATTACTAAAAAGATAATCAAGGTATGATTTACGCGATGTTTCAGAAATACATGATTA  
CTAAGGAAGAGATTTCCTTGAAAGAGCTTGATTTTAAATTGAGAAAACAACCTATTGAAAAACATAATCTTTACGGTAACATGGG  
TTCAGGAACAATCGTTATTAAAAATGAAAACGGTGGGAATATACGTTTGAATTACACAAAAAAGTCAAGAGCATCGTATGGCA  
30 GGCATAATATTGATAACATTGAAGTGAATATAAAATAA

SEQ ID No:88 - SSL3

35 gene 1..1059  
/gene="COLss13"  
CDS 1..1059  
/gene="COLss13"  
/product="(Similar to SET8 from strain N315)SSL2"  
/translation="MKMRTIAKTSIALGLLTGAI TVTTQSVKAEKIQSTKVVDKVP TLKAERLAMINITAGANSAT  
40 TQAANTRQERTPKLEKAPNTNEEKTSASKIEKISQPKQEEQKTLNISATPAPKQEQSQTTTESTTPKTKVTPFPST  
NTPQPMQSTKSDTPQSPTIKQAQTDMPKYEDLRAYYTKPSFEFEKQFGFMLKPWTTVRFMNVIPNRFYKIALVSG  
KDEKKYKDGPDNIDVFI VLEDNKYQLKKYSVGGITKTNSKKVNHKVELSITKKDNQGMISRDVSEYMITKEEISL  
45 KELDFELRKLQ LIEKHNLGNGSGTIVIKMNGGKYTFELHKKLQEHMAGTNIDNIEVNIK"

SEQ ID No:89

50 atgaaaataacaacgattgctaaaacaagtttagcactaggccttttaacaacaggtgtaatcacaacgacaacgc  
aagcagcaaacgcgacaacacccatcttccactaaagtgggaagcaccacaatcaacaccgcccctcaactaaaataga  
agcaccgcaatcaaaaccaaaccgcgacaacaccgcccctcaactaaagtgaagcaccgcaacaaacagcaaatgcg  
acaacaccgccttcaactaaagtgaacaacacccatccaacaaacacgcccacaaccaatgcaatctactaaatcag  
55 acacaccacaatcgccaaccacaaaacaagtagcaacagaaaataaaatcctaaatttaaagatttaagagcggtatta  
tacgaaaccaagtttagaatttaaaaatgagattggtattatttttaaaaaaatggacgacaataagattttatgaat  
gttgtccagattatttcatatataaaattgcttttagttggttaagatgataaaaaatatggtgaaggagtagacata  
ggaatgtcgatgtatttgcgttttagaagaaaaataattacaatctggaaaaatatctgtcggtggtatcacaaa  
gagtaatatgtaaaaaagttgatcacaaagcaggagtaagaattactaaggaagataataaaggtacaatctctcat  
60 gatgtttcagaattcaagattactaagaacagatttcttgaagaacttgatttttaattgagaaaaacaactta  
ttgaaaaaataatctgtacggtaacggttggttcaggtaaaattgttattaaaatgaaaaacggtggaaagtacac  
gtttgaattgcacaaaaaattacaagaaaatcgcatggcagatgtcattaatagtgaacaaattaaaaacatcgaa  
gtgaatttgaaa

65 SEQ ID No:90 - SSL4

gene 1..924  
/gene="COLss14"

CDS 1..924  
 /gene="COLss14"  
 /product="(Similar to SET9 from strain N315) SSL2"  
 5 /translation="MKITTTIAKTSALGLLTTGVITTTTQANATTPSSTKVEAPQSTPPSTKIEAPQSKPNATTP  
 PSTKVEAPQQTANATTPSTKVTTTPSTNTPOPMQSTKSDTPQSPSTTKQVPTEINPKFKDLRAYTKPSLEFKNEI  
 GIILKKWTTIRFMNVVPDYFIYKIALVGKDDKKYGEVHRNVDVFFVLEENNYNLEKYSVGGITKSNKKVDHKAG  
 VRITKEDNKGITISHDVSEFKITKEQISLKELDKFLRKQLIEKNNLYGNVSGKIVIKMKNNGGKYTFELHKKLQENR  
 10 MADVINSEQIKNIEVNLK"

SEQ ID No:91

15 atgaaattaacaacgatagctaaagcaacattagcattaggaatattaactacaggtgtgtttacagcagaaagtc  
 aaactgggtcacgcgaaagtagaacttgatgagacacaaacgcaaatattatcaaatatgctacatcaatactattc  
 tgaagaaagttttgaaccaacaaacattagtggttaaagcggaagattactatggctctaactgttttaacttttaa  
 caacgaaataaagcttttaagttttttacttggtgacgataaaaaataataaagaaaaaacacatggccttg  
 atgtctttgcagtcacctgaattaatagatataaaaggtggcatatatacggttggcggtataacaaagaaaaatgt  
 gagatcagtggtttgatttgaagtaatacgaagttcaaaaggttgatgctaaaaatggcttttcgata  
 20 aacgagttgtttttattcaaaaggaagaagatcattgaaggaactggacttttaaaataagaaaaactcttaatcg  
 aaaaatatagattgtataaaggaacgtctgataaaggtagaattgttatcaatatgaaagacgaaaagaagcatga  
 aattgatttaagtgaataaagttttgaacgtatgtttgatgtaattgtagtaagcaaattaaaaatattgaa  
 gtgaatttgaat

SEQ ID No:92 - SSL9

30 gene 1..696  
 /gene="COLss19"  
 CDS 1..696  
 /gene="COLss19"  
 /product="(Similar to SET13 from strain N315) SSL9"  
 35 /translation="MKLTIIAKATLALGILTGTGVTFAESQTGHAKVELDETQRKYINMLHQYYSEESFEPTNISV  
 KSEDIYGSNVLNFKQRNKAFKVFLLGDDKNKYKEKTHGLDVFAVPELIDIKGGIYVGGITKKNVRSVGFVSNPS  
 LQVKKVDAKNGFSINELFFIQKEEVSLKELDFKIRKLLIEKYRLYKGTSDKGRIVINMKDEKKHEIDLSEKLSFER  
 MFDVMSKQIKNIEVNLN"

SEQ ID No:93

40 Atgaaatttacagcattagcaaaagcgacattagcttttaggaattttaacaacaggaactttaacaacagaagttc  
 attcaggtcatgcaaaacaaaatcaaaagtcagtaataaacatgacaaggaagcattataccgatactacactgg  
 45 aaagactatggaaatgaaaaatattagtgctttgaaacatggtaaaaacaacttacgttttaagtttagaggtatt  
 aagattcaagttttactgcctggaaatgataaaagtaaatccaacagcgtagttatgaggggttagatgtttct  
 ttgttcaagaaaaaagagataagcacgatataatttttactgttggtggtgtaatacagaataataaaacatctgg  
 agttgtcagtgacccaatattaataattttcaaaagaaaaggtgaagatgcttttgtgaaaggttacccttattac  
 attaaaaaagaaaaataacactaaaagaactggattataagttgagaaagcatctaattgaaaaatacggacttt  
 50 ataaaaaatctcaaaagatggtagggtcaaaattagcttgaaagatggcagtttttataaccttgatttaagatc  
 taaattaaaaatttaatatatgggggaagtcatagaagcaacaaattaaagatattgaagttaacttaaag

SEQ ID No:94 - SLL10

55 gene 1..681  
 /gene="COLss110"  
 CDS 1..681  
 /gene="COLss110"  
 60 /product="(Similar to SET14 from strain N315) SSL10"  
 /translation="MKFTALAKATLALGILTGTTLTTEVHSGHAKQNQKSVNKHDKALRYRYTGKTMEMKNISAL  
 KHGKNNLRFRKFRGIKIQVLLPGNDKSKFQQRSYEGLDVFFVQEKRDKHDIFYTVGGVIQNNKTSQVVSAPILNISK  
 65 EKGEDAFVKGYPIYIKKEKITLKELDYKLRKHLEIKYGLYKTISKDGRVKISLKDGSFYNLDLRSKLFKFKYMGEVI  
 ESKQIKDIEVNLK"

SEQ ID No:95

5 atgaaattaaaaaatattgctaaagcaagtttagcactagggattttaacaacagggatgattacaactactgctc  
 agccagtaaaagcaagtacattagaggttagatcacaaagctactcaagacttgagtgaatattataatagaccgtt  
 ctttgagtatacaaatcagtcaggatataaaagaggaaggaagaaagtgacgtttactcctaattatcaacttatagat  
 gtaactttaactgggaatgaaaagcaaaattttgggtgaagatatttctaagttagatatatttgggtgaagagaaa  
 attctgatagatctggtaatacagcttcaattgggtggtattactaaaacaaacgggttcaaattatattgataaagt  
 10 aaaagatgtaaatttaataattactaaaaacatcgatagtggttacatcaacgtcaacatcatctacatatacaatt  
 aataaagaagaaatttcattaaaagaacttgatttttaattaagaaagcatttaattgataaacataacctttata  
 agacagaacctaaagacagtaaaattcgaattactatgaaagatgggtgggttctacacatttgaattgaataaaaa  
 gttacaaaacacaccgtatgggtgatgttattgatggcagaaatatagaaaaaattgaagtgaattta

15

SEQ ID No:96 - SSL11

gene 1..675  
 /gene="COLss11"  
 20 CDS 1..675  
 /gene="COLss11"  
 /product="(Similar to SET15 from strain N315)SSL11"  
 25 /translation="MKLNIAKASLALGILTTGMITTTAQPVKASTLEVRSQATQDLSEYYNRPFFEYTNQSGYKE  
 EGKVTFTPNYQLIDVTLTGNEKQNFGEISNVDIFVVRENSDRSGNTASIGGITKTNGSNYIDKVKDVNLIITKNI  
 DSVTSTSTSSYITINKEEISLKELDFKLRLKHLIDKHNLYKTEPKDSKIRITMKDGGFYTFELNKKLQTHRMGDVID  
 GRNIEKIEVNL"

30 SEQ ID No:97

atgaaaaagaacatcatgaataaattagttttatcaacagcattgttacttttagaaactacatcaacacaacttc  
 ctaaaacaccaatcagtttttcatctgaagcaaaagcctataatatcagtgaaaacgagactaatatcaatgaact  
 35 aatcaaatattacactcagccgcatTTTTcattatctggaaaatgggtatggcaaaagcccaatggtagcattcat  
 gcaacattgcaaacgtgggttttggtatagtcattcaagtgtttggatccgagagttggggaaacattaatcagt  
 taagaaataaatacgttgatataatttggaactaaagatgaggacacagttgaagggttactggacttatgatgaaac  
 atttactgggtggtgttacgccagcagctacttcatctgataagccttatagactatTTTTTaaatatagtgataaa  
 caacaaactatcatcggtggacatgaattttacaaaggaataaaccagttattaacttttaaaagaattagatttcc  
 gtattcgtcaaacatttaataaaaaataaaaaagttatataacggagaatttaataaagggtcaaatgaataaactgc  
 40 tgatggaaataattacacgattgatttaagtaaaaagttaaaattaactgacacaaaccggttatgttaaaaatcct  
 cgtaatgcagaaattgaagtcatactcgaaaaatctaac

45 SEQ ID No:98 - SSL12

gene 1..723  
 /gene="COLss12"  
 CDS 1..723  
 /gene="COLss12"  
 50 /product="(similar to SA1011)SSL12"  
 /translation="MKKNIMNKLVLSTALLLLETTSTQLPKTPISFSSEAKAYNISENETNINELIKYYTQPHF  
 SLSGKWLWQKPNQSIHATLQTVVWYSHIQVFGSESWGNINQLRNKYVDIFGTDKEDTVEGYWYDETFTGGVTPAA  
 TSSDKPYRLFLLKYSQKQTTIIGHEFYKGNKPVLTLLKELDFRIRQTLIKNKKLYNGEFNKGQIKITADGNNTYIDL  
 55 SKKLKLTDTNRYVKNPRNAEIEVILEKSN"

55

SEQ ID No:99

60 atgaacaataacatcacgaaaaaattatTTTTatcaacaacattgttactattaggtacagcatctacacaatttc  
 ctaatacacctatcaattcttcatctgaagcgaaagcttattatataaatcaaaacgaaactaacgttaattgagtt  
 aactaaatattactcgaaaaatatttaaccttctctaacagtcagttatggcaaaaagataacggtagcattcat  
 gcaacgttggttacagttttcttggtatagtcattcaagtttatggacctgaaagttggggcaatatcaaccaat  
 taagaaataaaagcgttgatatttttggcataaaagaccaagaaccattgattcttttgcatattctcaagaaac  
 65 gtttactgggtggtgttactcctgcagcaacatctaacgataaaactgaatgtaacatatataagataaa

gcagaaacggtttactggcggattttccagtttatgaaggcaataagcctgttttaacttttaaagaattagatttttc  
gtattcggtcaaacatttaattaaaagtaaaaaattatataataattccttataataaaggacaaattaaaaataacagg  
tgcagacaataactacacaatagattttaagtaaaagggttgccatcaactgatgcaaatagatatgttaaaaaacct  
caaaatgcaaaaattgaagttatcctcgaaaaatcaaac

5

SEQ ID No:100 - SSL13

10      gene                      1048124.. 1048849  
                                 /gene="COLss13"  
                                 1048124.. 1048849  
                                 /gene="COLss113"  
                                 /product="(similar to SA1010 from strain N315)SSL13"  
15      /translation="MNNNITKKIILSTLLLLGTASTQFPNTPINSSSEAKAYYINQNETNVNELTKYYSQKYL  
         TFSNSTLWQKDNGTIATLLQFSWYSHIQVYGPESWGNINQLRNKSVDFIGIKDQETIDSFALSQETFTGGVTPAA  
         TSNDKHYKLVNTYKDKAETFTGGFPVYEGNKPVLTLLKELDFRIRQTLIKSKKLYNNSYNKGQIKITGADNNYTIDL  
         SKRLPSTDANRYVKKPQNAKIEVILEKSN"

20      SEQ ID No:101

atgagtaagaacatcacgaaaaatataattttaacgacaacattattactattaggtactgtattacctcaaaatc  
aaaaaccagttatttagtttttactctgaagctaaagccttatagcattgggtcaagatgaaactaacatcaatgaatt  
aattaaatatttacacacagcctcatttttccattttcaaataaatggctatatcaatatgataatggaaacattttat  
25      gttgaacttaagagatattcatgggtcagcacatatatctttatggggcgctgaaagttgggggaaatattaatcagt  
         taaaagatcggttacgtagatgtgtttggactaaaagacaaagatactgatcagttatgggtggtccttatagagagac  
         atttacaggtggcggttacaccagccgcaaaaccttctgataaaaacttataatctttttgtgcaatacaaaagataaa  
         ctacaaacgattattgggtgcgcatataaaatataccaaggcaataaaccagttattaacattgaaagaaatcgatttcc  
         gtgcacgagaagcggttaataaaaaataaaatattatataacgaaaatcgtaataaaggtaagcttaagatcaccgg  
30      tggcggttaataactacactattgatttaagcaaaagattacattcagatctagcaaatgtttatgttaaaaaatcct  
         aataaaataactgttgacgtcctctttgat

SEQ ID No:102 - SSL14

35      gene                      714.. 714  
                                 /gene="COLss14"  
                                 714.. 714  
                                 /gene="COLss114"  
                                 /product="(similar to SA1009 from strain N315)SSL14"  
40      /translation="MSKNITKNIIILTTLLLLGTVPQNQKPVFSFYSEAKAYSIGQDETNNINELIKYYPQPHF  
         SFSNKWLYQYDNGNIYVELKRYSWSAHISLWGAESWGNINQLKDRYVDVFLGKDKDQDLWWSYRETFTGGVTPAA  
         KPSDKTYNLFVQYKDKLQTIIGAHKIYQGNKPVLTLEIDFRAREALIKNKILYNENRNKGKLIKITGGGNNYTIDL  
         SKRLHSDLANVYVKNPNKITVDVLFDF"

45      SEQ ID No: 103

*S. aureus* strain N315 (SSL12-SSL14) -coding strand

50      1      C T A C T G T T C T   T T G T G A A A A A   C C A C G G T A T T   C A A T G C C A T C   A T A C A T T C C A   C C A A G C A C A C  
61      G T G C A G T A T C   T T T A G T T G T T   T C T T T T T T A C   C C A T T T G T G A   T C C A G T T G G G   C C T A A A T A A G  
121      T T A C A T T T G C   A C C T T G A T C A   T G C G C T G C A A   C T T C A A A T G C   A C A T C G C G T T   C T T G T A G A A T  
181      C T T T T T C A A A   T A A C A G T G C A   A T A T T T T T A T   T T T T T A A C A T   A G G C T T T T C A   G T G C C A A T A T  
241      A T T T A G C A C G   T T T T A A A T C C   T C G G A G A G T G   T T A A T A A G A A   T T C T A C C T C T   T G T C G T G A A A  
55      301      A G T C T A A T A A   A G T T A A A A A A   C T T C T G T T T C   G T A A A T T T T T   C A T T A A T A A C   A A C T C C T T T T  
361      A T T T T A G T T A   C A C C T A A A T T   G T A A G C G C A T   A C A A A T A A A A   C A C A A C C T A C   T A T T A A A A T T  
421      T G T A A T A T T T   T A T C A A T A A T   T A A A T G A A C A   T T T T A T T A A T   A T T A A A T T T A   A G T A G T A G G A  
481      A A T A A T T A A A   A T A A G T A C T A   C A T T T A A A G T   A T A A C T A T T T   T T C A A G T A G T   T A G A A A A T T C  
541      A A T T A T C A A A   C A A T T T A A T G   C A A T T G A T T A   G A G A A T A A T T   G T A A C G T G T C   G T T T T T A A T A  
60      601      T A T A A C T C C C   G C C T A C T T T A   T T A A G T A C T G   T T T C T G T C C A   A A A C T T A A A A   A T G A T A A G T T  
661      T T G C T T A A A T   A A C A C T A C T A   A C T G T T T A A G   T T T A T T T A A C   A T A G T T T T A G   C T T T T A T T T A  
721      A T T C C G A A T C   G G T G T A A T A G   C T T A T A T A C T   T T G G G T A A T T   C A C G C A A A G G   A G A T T T T C A T  
781      A T G A A A A A G A   A C A T C A T G A A   T A A A T T A G T T   T T A T C A A C A G   C A T T G T T A C T   T T A G G A A C C  
841      A C A T C A A C A C   A A C T T C C T A A   A A C A C C A A T C   A G T T T T T C A T   C T G A A G C A A A   A G C C T A T A A T  
65      901      A T C A G T G A A A   A C G A G A C T A A   T A T C A A T G A G   T T A A T C A A A T   A T T A C A C T C A   G C C G C A T T T T  
961      T C A T T A T C T G   G A A A A T G G C T   A T G G C A A A A G   C C C A A T G G T A   G C A T T C A T G C   A A C A T T G C A A  
1021      A C G T G G G T T T   G G T A T A G T C A   T A T T C A A G T G   T T T G G A T C C G   A G A G T T G G G G   A A A C A T T A A T

1081 CAGTTAAGAA ATAAATACGT TGATATATTT GGAACATAAG ATGAGGACAC AGTGAAGGT  
 1141 TACTGGACTT ATGATGAAAC ATTTACTGGT GGTGTTACGC CAGCAGCTAC TTCATCTGAT  
 1201 AAACCTTATA GACTATTTTT AAAATATAGT GATAACAAC AACTATCAT CGGTGGACAT  
 1261 GAATTTTACA AAGGAAATAA ACCAGTATTA ACTTTAAAG AATTAGATTT CCGTATTCGT  
 5 1321 CAAACATTAA TAAAGAATAA AAAGTTATAT AACGGAGAAT TTAATAAAGG TCAAATTAAG  
 1381 ATAACCTGCTG ATGGAAATAA TTACACGATT GATTTAAGTA AAAAGTTAAA ATTAACGTAC  
 1441 ACAAACCGTT ATGTTAAAAA TCCTAAAAAT GCACAAATG AAGTCATAC CGAAAAATCT  
 1501 AACTAACCTA TTACCTTTTG TAAATGCGGA TAATTTCAAT TATCTAATTA ACCCCTTTTT  
 1561 ATAATTAAC ATTCCAACAA TACTCAAAGG AGAAATTCGA ATGAACAATA ACATCACGAA  
 10 1621 AAAAATTATT TTATCAACAA CATTGTTACT ATTAGGTACA GCATTTACAC AATTCCTTAA  
 1681 TACACCTATC AATTCTTCAT CTGAAGCGAA AGCTTATTAT ATAAATCAAA ACGAAACTAA  
 1741 CGTTAATGAG TTAACATAAT ATTACTCGCA AAAATATTTA ACCTTCTCTA ACAGTACGTT  
 1801 ATGGCAAAAA GATAACGGTA CGATTCATGC AACGTTGTTA CAGTTTCTT GGTATAGTCA  
 1861 TATTCAAGTT TATGGACCTG AAAGTTGGGG CAATATCAAC CAATTAAGAA ATAAAGCGT  
 15 1921 TGATATTTTT GGCATAAAAG ACCAAGAAAC CATTGATTCT TTTGCATTAT CTCAAGAAAC  
 1981 GTTTACTGGT GGTGTTACTC CTGCAGCAAC ATCTAACGAT AAACACTATA AACTGAATGT  
 2041 AACATATAAA GATAAAGCAG AAACGTTTAC TGGCGGATTT CCAGTTTATG AAGGCAATAA  
 2101 GCCTGTTTTA ACTTTAAAG AATTAGATTT TCGTATTCTG CAAACATTAA TTAAGGTAA  
 2161 AAAATTATAT AATAATCTCT ATAATAAAGG ACAAATTAAT ATAACAGGTA CAGACAATAA  
 20 2221 CTACACAATA GATTTAAGTA AAAGGTTGCC ATCAACTGAT GCAAATAGAT ATGTTAAAAA  
 2281 ACCTCAAAAT GCAAAAATTG AAGTTATCCT CGAAAAATCA AACTAACAT AATAATGGAG  
 2341 TTAATAAAAA TAATCGCAAA TACTATATTG ACTTCGCTCA CATTTAAAT TCTTATTCCT  
 2401 CGTATCATGA TTCCTCTGAA AGGAGATGTT CTAATGAGTA AGAACATCAC GAAAAATATA  
 2461 ATTTTAACGA CAACATTATT ACTATTAGGT ACTGTATTAC CTCAAAATCA AAAACAGTA  
 25 2521 TTTAGTTTTT ACTCTGAAGC TAAAGCTTAT AGCATTGGTC AAGATGAAAC TAACATCAAT  
 2581 GAATTAATTA AATATTACAC ACAGCCTCAT TTTTCATTT CAAATAAATG GCTATATCAA  
 2641 TATGATAATG GAAACATTTA TGTTGAACCT AAGAGATATT CATGGTCAGC ACATATATCT  
 2701 TTATGGGGCG CTGAAAGTTG GGGAAATATT AATCAGTTAA AAGATCGTTA CGTAGATGTG  
 2761 TTTGGACTAA AAGACAAAAG TACTGATCAG TTATGGTGGT CTTATAGAGA GACATTTACA  
 30 2821 GGTGGCGTTA CACCAGCCGC AAAACCTTCT GATAAACTT ATAATCTTTT TGTGCAATAC  
 2881 AAAGATAAAC TACAAACGAT TATTGGTGGC CATAAAATAT ACCAAGGCAA TAAACAGTA  
 2941 TTAACATTGA AAGAAATCGA TTTCCGTGCA CGAGAAGCGT TAATAAAAAA TAAATATTA  
 3001 TATACCGAAA ATCGTAATAA AGGTAAGCTT AAGATCACCG GTGGCGGTAA TAACACACT  
 3061 ATTGATTTAA GCAAAAGATT ACATTCAGAT CTAGCAAATG TTTATGTTAA AAATCCTAAT  
 35 3121 AAAATAACTG TTGACGTCTT CTTTGATTAG TATATGAAGG TGACTTATAC TTCATGCACT

SEQ ID No. 104

*S. aureus* strain NCTC8325 (SSL1-SSL11) – coding strand

40 1 TTGATTAAAA TAATCTAATT GTCGAACAGG AACTTTTCCG CGCCAATCTT CTGGAACTTT  
 61 CTCAACATTT CTAACACCAA TGTGAAAATG ATCTATGTGA TTTGCAATGG CTTGATTGTG  
 121 AATATGTGTG CCTAAATGAC CTGTAGCACC TGTTAAACATA ATATTCATTC ACTTCATCTC  
 181 CTAATCTTTA TATACATAAC ATAATACTTA TTTGATGGTT TTCAAAACAT TTGATTTTAT  
 45 241 AAAAAATTTCT AATCTGTATT TATTGTGCGC GTGTATAGTA AATACGTAAA TATTATTAAT  
 301 GTTGAAGGATG CCGTAATGAC GCGTTTTAGT TGATGTGTAT CACTAATATC ATTGAAAATT  
 361 TTAATCAGGT ACTACGACAA TATGATGTCT GTTTTGTGTC TGAAAGTTTT ACAGTTTTTA  
 421 AAATAAAAAAT GGTATAAAGT GTGATTTGTA TAAAAAGAG TCTCGACGGA TAAGAATTGA  
 481 TTAATAACAG TTAGCATTTT ATTAATTACC TTAACAATGA TTCAAGTTTA GTTAAATGAG  
 50 541 GTTTAATTTG AAAGGGGATA GCGCCTCAAT ATAATGTAGG TAGATTGTTT ATATTACGTA  
 601 ATTGAAAAAT CAAATTTAA TAGATTGGGG CTAATAATTA TGAAATTTAA AGCGATAGCA  
 661 AAAGCAAGTT TAGCATTTGG AATGTTAGCA ACAGGTGTAA TTACATCGAA TGTACAATCA  
 721 GTACAAGCGA AAGCAGAAGT TAAACAACAA AGTGAATCAG AGTTAAACA CTATTATAAT  
 781 AAACCAATTT TAGAGCGTAA AAATGTGACT GGATTTAAAT ATACTGATGA GGGTAAACAC  
 55 841 TATTTAGAAG TCACAGTAGG GCAACAGCAT TCTCGAATCA CTTTACTTGG ATCTGATAAA  
 901 GATAAATTTA AAGACGGAGA AAACCTCAAT ATAGATGTGT TTATCCTTAG AGAAGGTGAC  
 961 AGTAGACAAG CAACAAATTA CTCAATTGGT GCGGTTACAA AATCAAATAG TGTGCAGTAT  
 1021 ATTTGATTTA TCAATACGCC AATTTTAGAA ATCAAGAAAG ATAATGAAGA TGTACTTAAA  
 1081 GATTTTACT ACATTTCAAA AGAAGACATC TCATTAAAG AACTTGATTA TAGATTAAGA  
 60 1141 GAACGTGCGA TTAAACAACA CGGCTTGTAT TCAATGGTTC TTAACAAGG TCAAATTACA  
 1201 ATTACAATGA ATGATGGCAC AACACATACA ATCGATTAA GTCAAAAAC TGAAGAAAGAA  
 1261 CGTATGGGTG AGTCAATCGA CGGCACCTAG ATTAATAAAA TTCTAGTAGA AATGAAATAA  
 1321 TACTTTCTAA CAACAAGCGG CTATGTTGAA TAGTGCTTGT TATGGAAATA TATGGAAGTT  
 1381 AAGCGACGTA CTGTTGCTTA GCTTCTTTTT TTGAGGGGAA AAGTTACAAA ACTCACACAA  
 65 1441 ACAGTCGCAC CACGCATTAT CTTTTGCTTA AATAGCTTAA TCATATTTTA TGAATAGTTA  
 1501 AAAACAGGTT AATGTGAATA TCCGAATACA GTCCTATAA TATGGGTGTA TGATTCAAAT  
 1561 TACGTAATAA AACAATCTAA TTATAATAGA TTGGAGCATA CAACTATGAA AATGAAAAAT  
 1621 ATTGCAAAAA TAAGTTTGTT ATTAGGAATA TTAGCAACAG GTGTAAACAC TACACGGAA  
 1681 AAACCAAGTT ATGCCGAAA GAAACCTATT GTAATAAGTG AAAATAGCAA AAAATTAATA  
 70 1741 GCTTATTATA ATCAACCTAG TATTGAATAT AAAAATGTGA CAGGTTATAT CAGTTTCATT  
 1801 CAACCAAGTA TTAATTTTAT GAATATCATA GATGGTAATT CTGTTAATAA TATTGCTTTA  
 1861 ATTGGCAAG ATAAGCAACA TTATCATACG GGTGTACATC GTAATCTTAA TATATTTTAC  
 1921 GTTAATGAGG ATAAGAGATT TGAAGGTGCA AAGTACTCTA TTGGGGGTAT CACGAGTGCA

1981	AACGATAAAG	CTGTCGACCT	AATAGCAGAA	GCAAGAGTTA	TTAAAGAAGA	TCATACTGGT
2041	GAATATGATT	ATGACTTTTT	CCCATTAAAA	ATAGATAAAG	AAGCGATGTC	ATTGAAAGAG
2101	ATTGATTTTA	AATTAAGAAA	ATACCTTATT	GATAATTATG	GTCTTTACGG	TGAAATGAGT
2161	ACAGGAAAAA	TTACAGTCAA	AAAGAAATAC	TATGGAAAGT	ATACATTGA	ATTGGATAAA
2221	AAGTTACAAG	AAGACCGTAT	GTCCTGATGT	ATCAATGTCA	CAGATATTGA	TAGAATTGAA
2281	ATCAAAGTTA	TAAAGCATA	ACACATATAC	TTGATGACGA	AATAAGTTGA	AATTGAAATA
2341	GAGAGGTTAA	GTGACGATCA	AACGTTGCTT	AACCTCTTTT	TAATGCTTAA	AAATTATTC
2401	AAAGGCACAT	AGAAACGCTA	TATTAATCTC	ATACTCACTC	ATTATTTTTT	GCTTAAATTA
2461	CTTAATAATA	CTTCAATAAT	TGTTAAAAGG	GGTTTAAATG	GATTATCTTA	GAACGCCATC
2521	TATAATGATG	TTGTATGATT	CAAATTACGT	AAAAAGACAA	TGGAATATAA	TATAGATTGG
2581	AGCATACAAT	TATGAAAATG	AGAACAATTG	CTAAAACCAG	TTTAGCCTTA	GGGCTTTTAA
2641	CAACAGGCGC	AATTACAGTA	ACGACGCAAT	CGGTCAAAGC	AGAAAAAATA	CAATCAACTA
2701	AAGTTGACAA	AGTACCAACG	CTTAAGACAG	AGCGATTAGC	AATGATAAAC	ATAACAGCAG
2761	GTGCAAAATC	AGCGACAACA	CAAGCAGCTA	ACACAAGACA	AGAAGCGCAG	CCTAAACTCG
2821	AAAAGGCACC	AAATACTAAT	GAGGAAAAAA	CCTCAGCTTC	CAAAATAGAA	AAAATATCAC
2881	AACCTAAACA	AGAAGAGCAG	AAAACGCTTA	ATATATCAGC	AACGCCAGCG	CCTAAACAG
2941	AACAATCACA	AACGACAACC	GAATCCACAA	CGCCGAAAAC	TAAAGTGACA	ACACCTCCAT
3001	CAACAAACAC	GCCACAACCA	ATGCAATCTA	CTAAATCAGA	CACACCACAA	TCTCCAACCA
3061	TAAAACAAGC	ACAAACAGAT	ATGACTCCTA	AATATGAAGA	TTTAAGAGCG	TATTATACAA
3121	AACCGAGTTT	TGAATTTGAA	AAGCAGTTTG	GATTTATGCT	CAAAACCATGG	ACGACGGTTA
3181	GGTTTATGAA	TGTTATTCCA	AATAGGTTCA	TCTATAAAAT	AGCTTTAGTT	GGAAAAGATG
3241	AGAAAAATA	TAAAGATGGA	CCTTACGATA	ATATCGATGT	ATTTATCGTT	TTAGAAGACA
3301	ATAAATATCA	ATTGAAAAAA	TATTCTGTCTG	GTGGCATCAC	GAAGACTAAT	AGTAAAAAAG
3361	TTAATCACA	AGTAGAATTA	AGCATTACTA	AAAAAGATAA	TCAAGGTATG	ATTTACCGCG
3421	ATGTTTTCAGA	ATACATGATT	ACTAAGGAAG	AGATTTCTTT	GAAAGAGCTT	GATTTTAAAT
3481	TGAGAAAACA	ACTTATTGAA	AAACATAATC	TTTACGGTAA	CATGGGTTCA	GGAACAATCG
3541	TTATTAAAAAT	GAAAAACGGT	GGGAAATATA	CGTTTGAATT	ACACAAAAAA	CTGCAAGAGC
3601	ATCGTATGGC	AGACGTCATA	GATGGCCTA	ATATTGATAA	CATTGAAGTG	AATATAAAAT
3661	AATCATGACA	TTCTCTAAAT	AGAAGCTGTC	ATCGGAAAAA	CAAGAAAGTTA	AGTGACACAG
3721	GTTTACATGT	TGCTTAGCTT	CTTTTATTAT	CGCTAATGAT	GTA AAAAGAC	GAATATTTCAT
3781	TTGTTTGTAA	AAGTGGCATT	TCTATGTCTT	AAAAGTGACG	AAACTTCAAA	TGTGCCAAGT
3841	TTGTAATCAG	ATCAAAATCA	TTTTTATTTA	ACGAACATTA	TGGATTCTTT	AATTTACTTA
3901	ACGATGATTC	AAATATAGTT	AAACAAGGTT	TAATGTGAAT	GGAGCAATAC	GCCATCTATA
3961	ATAAAGCTGT	ATGATTCAAT	GAATGTAATC	GAACAAATCT	AATAATTACG	AATGGAGCAT
4021	ACAACATATGA	AAATAACAAC	GATTGCTAAA	ACAAGTTTAG	CACTAGGCCT	TTTAACAACA
4081	GGTGTATCA	CAACGACAAC	CAACGACAGCA	AACGCGACAA	CACTATCTTC	CACTAAAGTG
4141	GAAGCACCAC	AATCAACACC	GCCCTCAACT	AAAATAGAAG	CACCGCAATC	AAAACCAAAC
4201	CGGACAACAC	CGCCCTCAAC	TAAAGTAGAA	GCACCGCAAC	AAACAGCAAA	TGCGACAACA
4261	CCGCCCTTCA	CTAAAGTGAC	AACACCTCCA	TCAACAAACC	CGCCACAACC	AATGCAATCT
4321	ACTAAATCAG	ACACACCACA	ATCGCCAACC	ACAAAACAAG	TACCAACAGA	AATAAATCCT
4381	AAATTTAAAG	ATTTAAGAGC	GTATTATACG	AAACCAAGTT	TAGAATTTAA	AAATGAGATT
4441	GGTATTATTT	TAAAAAATG	GACGACAATA	AGATTTATGA	ATGTTGTCCC	AGATTATTTT
4501	ATATATAAAA	TGCTTTTAGT	TGGTAAAGAT	GATAAAAAAT	ATGGTGAAGG	AGTACATAGG
4561	AATGTCGATG	TATTTGTCTG	TTTAGAAGAA	AATAATTACA	ATCTGGAAAA	ATATTCTGTC
4621	GGTGGTATCA	CAAAGAGTAA	TAGTAAAAAA	GTTGATCACA	AAGCAGGAGT	AAGAATTACT
4681	AAGGAAGATA	ATAAAGGTAC	AATCTCTCAT	GATGTTTCAG	AATTCAAGAT	TACTAAAGAA
4741	CAGATTTTCT	TGAAAGAACT	TGATTTTAAA	TTGAGAAAAC	AACCTATTGA	AAAAAATAAT
4801	CTGTACGGTA	ACGTTGGTTC	AGGTAAAATT	GTTATTAAAA	TGAAAAACGG	TGGAAAGTAC
4861	ACGTTTGAAT	TGCACAAAAA	ATTACAAGAA	AATCGCATGG	CAGATGTCAT	AGATGGCACT
4921	AATATTGATA	ACATTGAAAT	TAAATATAAA	TAATCATGAC	ATTCTCTAAA	TAGAAGCTGT
4981	CATCGGAAAA	ACAAGAAGTT	AAGTGACAAC	GGCCTACATG	TTGCTTAGCT	TCTTTTGTTA
5041	TGTTTCGATGA	TTTGAGAAC	CGAATTTTCG	ATGGGTCCAA	ATATGACGTG	GAAGAGACCT
5101	GAATTTATCT	GTAATCCCT	ATCTATCGGG	TGTGAAGCAC	AACGGGATCA	GTTTTATTTA
5161	ACGAACATTA	TAGATTCCCT	AATTTACTTA	ATAATGATTC	AATGATTATT	AAACATGGTT
5221	TAATGTGAAA	GGTCAAATAC	GCTAACTATA	ATAAAGCTGT	ATGATTCAAT	AGACGTAAGC
5281	GAACAAATCT	ATAAATTACG	AATGGAGCAT	ACAACTATGA	AAATGACAGC	AATTGCCGAA
5341	GCAAGTTTAG	CATTAGGTAT	TTTAGCAACA	GGAACAATAA	CGTCATTGCA	TCAAACGTGA
5401	AATGCGAGTG	AACATAAAGC	AAAATATGAA	AATGTGACAA	AAGATATCTT	TGACTTAAGA
5461	GATTACTATA	GTGGCGCAAG	TAAGGAACCT	AAAAATGTTA	CTGGTTATCG	TTATAGCAAA
5521	GGTGGCAAGC	ATTACCTTAT	CTTTGATAAA	AATAGAAAAA	TCACAAGAGT	ACAGATATTT
5581	GGTAAAGATA	TTGAAAGATT	TAAAGCACGT	AAAAATCCGG	GATTAGACAT	ATTTGTTGTT
5641	AAAGAAGCGG	AAAACCGTAA	TGGCACAGTG	TTTTCATATG	GTGGTGTCTC	TAAGAAAAAT
5701	CAAGACGCTT	ATTATGATTA	TATAAACGCA	CCAAGATTTC	AAATCAAGAG	AGATGAAGGT
5761	GACGGTATTG	CTACGTACGG	TAGAGTACAC	TACATTATTA	AAGAAGAGAT	TTCACTTAAA
5821	GAACTCGACT	TTAAATTGAG	ACAGTATTTA	ATTCAAAAAT	TTGATCTGTA	TAAAAAGTTT
5881	CCTAAAGATA	GTAAGATAAA	AGTGATAATG	AAAGATGGCG	GCTATTATAC	GTTTGAACCT
5941	AATAAAAAAT	TACAAACAAA	TCGCATGAGT	GACGTCATTG	ACGGTAGAAA	TATTGAAAAA
6001	ATAGAAAGCA	ACATTAGATA	ATTCAATGAA	ATATGGATAA	TAGTAAAAATA	TGGATAGTAT
6061	AGAGGAGTTA	GGCAACATAA	GTTGCTTAGC	TTCTTTTTTG	TGTTGGAGAG	ATGAAAATGA
6121	AGCGTATCGA	TGAATAATTA	AAACACCAAT	AAAACCTTGT	GAAATAGTTG	ATACTTATAG
6181	TCGCGCGTTG	TCCTTTTCGT	GACATGAAAC	AATGTGGAAA	ACATAGTTAA	ATTGAGGGAA
6241	ACTGTGAATA	GTTAAAAAAG	CTCGTTTAA	TTTAAAAAAT	AGATTAACGC	TGTTAGGATT
6301	CCATTAATTA	GCTTAACATT	GGTCAAAAA	TAGTTAAAAA	GAGGTAAATT	CATAGCTTAG

5 6361 TATTACGCTT ATATAATGAT AGTAGATTGT TCGTATTACG TAATTGAAAT AATCATATAA  
 6421 AAATATATTA AGACAAAATT TATAAATAGA TTGGGAGAAT AGTACTATGA AATTAAAAAC  
 6481 GTTAGCTAAA GCAACATTAG TATTGGGATT GTTAGCTACT GGTGTAATAA CAACAGAAAG  
 6541 TCAAACAGTA AAAGCGGCAG AATCAACTCA AGGTCAACAC AATTATAAAT CGTTAAAAATA  
 6601 CTAATATAGC AAGCCAAGTA TAGAGTTAAA AAATCTTGAT GGTGTTGTATA GACAGAAAGT  
 6661 GACAGATAAA GGAGTATATG TTTGGAAGGA TCGAAAAGAT TATTTTGTG GCTTGCTTGG  
 6721 TAAAGATATT GAAAAATACC CTCAGGTGA GCATGATAAG CAAGATGCAT TTTTAGTCAT  
 6781 CGAGGAGGAA ACTGTTAATG GAAGACAATA TTCAATTGGT GGTGTTAAGTA AGACAAATAG  
 6841 TAAAGAATTT AGTAAAGAAG TCGATGTTAA AGTAACAAGA AAAATTGATG AATCATCGGA  
 10 6901 AAAGTCTAAA GATAGTAAAT TTAARATTAC TAAAGAAGAA ATCTCGTTAA AAGAGTTGGA  
 6961 CTTTAAATTA AGAAAAAAT TGATGGAAGA AGAAAAATTA TATGGTGCTG TTAATAATAG  
 7021 AAAAGGTAAA ATTGTAGTTA AAATGGAAGA TGATAAGTTT TATACTTCG AACTTACAAA  
 7081 AAAACTACAA CCGCATCGCA TGGGTGACAC GATAGATGGT ACCAAATCA AAGAAATTAA  
 15 7141 TGTTGAGCTA GAATATAAAT AATCTTTGGA CAAGCAGACT AGTAATTGTA GGGAAAGTTAA  
 7201 GCGATAACAT ATTGCTTAGC TTCTTTTTTA TTTTGTATG ATGAAAAAG GAGCGGGTTT  
 7261 ATGATCAAGT TTTTGGAAAA ACGGTGATA CTTATAGTCG CGCGTTGTCC TTTTCGTGAC  
 7321 ATGAAACAAAT GTGGAAAAACA TAATTAAATT GAGGGAAAGT GTGAATAGTT AAAAAATTAG  
 7381 TATTGTGTTA TAAAAAATAA TTAATACTGT TAGGATTTCA TTAACCTAAT TAACGTTGTT  
 7441 TCAAAAATAG TTA AAAAGAG GTTAATTCAT AGCGCAGTAT CTCATTATA TAATGATAGT  
 20 7501 AGATTGTTTCG TATTACGTAA TTGAATTAA CATATAAATA TATATTAGA CAAATTTATA  
 7561 AATAGATTGG GAGAATAGTA CTGTGAAATT AAAACGTTA GCTAAAGCAA CATGGGCATT  
 7621 AGGCTTATTA ACTACTGGTG TGATTACATC AGAAGGCCAA GCAGTTCAAG CAAAAGAAAA  
 7681 GCAAGAGAGA GTACAACATT TATATGATAT TAAAGACTTA CATCGATACT ACTCATCAGA  
 7741 AAGTTTTGAA TTGAGTGTAA GGTGAAAAAT TATAACGGTT CTAACGTTGT  
 25 7801 ACGCTTTAAC CAAGAAAATC AAAATCACCA ATTATTCTTA TTAGGTAAAG ATAAAGAGAA  
 7861 ATATAAGAA GGCATTGAAG GCAAAGATGT CTTTGTGGTA AAAGAATTAA TTGATCCAAA  
 7921 CGGTAGATTA TCTACTGTTG GTGGTGTGAC TAAGAAAAAT AACAAATCTT CTGAAACTAA  
 7981 TACACATTTA TTTGTTAATA AAGTGTATGG CGGAAATTTA GATGCATCAA TTGACTCATT  
 8041 TTCAATTAAT AAAGAAGAAG TTTTCACTGAA AGAACCTGAT TTCAAAATTA GACAACATTT  
 30 8101 AGTTAAAAAT TATGTTTTAT ATAAAGGTAC GACTAAATAC GGTAAAGTCA CTATCAATTT  
 8161 GAAAGATGGA GAAAAGCAAG AAATTGATTT AGGTGATAAA TTGCAATTCTG AGCGCATGGG  
 8221 TGATGTGTTG AATAGTAAGG ATATTATAAA GATTGAAGTG ACTTTGAAAC AAATTTAAAG  
 8281 TAAGTAATCA ATGACTCTAA AGTAATAAAT TTGAAGCAGC TTAACGATGA AATGTTGAAT  
 35 8341 AGATACGTTA ACCTTACATT AAGGAGCGTA TTTAAAACAA CCTTGTCGTT AGGCTTTTTT  
 8401 TTACGTTTTA TAACGCAGGT TATGAGTGAG CATACTAAAA ATTTACATTG CTCTTGAAG  
 8461 TGATGCCCAT TGAATATTAA TTAGCTCTTC ATTAACAAATG ATTTAAGTTT AATTAACGA  
 8521 GTGTTAATGT CAGTCTGTCT CAATGCCCTT TATAATAAAG GTGTATTATT CAAATTACGT  
 8581 AATAAAGCA ATCCAATATA TTAAGATTGG AGCACATGAA TATGAAATTT ACAGTGATAG  
 40 8641 CTAAAGCGAT ATTTATATTA GGAATATTAA CAACAAGTGT AATGATAACA GAAAATCAAT  
 8701 CGGTTAATGC AAAAGGAAAG TATGAAAAAA TGAACCGTTT ATATGATACA AACAAAGTTAC  
 8761 ATCAATACTA TTCAGGACCT AGTTATGAGT TAACAAATGT TAGTGCCCAA AGTCAAGGTT  
 8821 ATTTATGACT TAACGTTTTG CTTTTTAACC AACAAAATCA AAAGTTCCAA GTGTTTTTAT  
 8881 TGGGAAAAGA TGAAAATAAA TACAAAGAAA AAACACATGG TTTAGATGTC TTTGCGGTAC  
 8941 CAGAATTAGT AGATTTAGAT GGAAGAATAT TTAGTGTTAG TGGTGTAACA AAGAAAAACG  
 45 9001 TAAATCAAT ATTTGAGTCT CTAAGACGCG CGAACTTACT AGTTAAAAAA ATAGACGATA  
 9061 AAGACGGTTT TTCTATTGAT GAATTTTTCT TTATTCAAAA GGAAGAAGTG TCATTGAAGG  
 9121 AACCTGATTT TAAAAATAAG AAAGCTTTGA TTA AAAAATA CAAACTGTAT GAAGGGTCAG  
 9181 CTGATAAAGG TAGAATTGTT ATTAATATGA AAGATGAAAA TAAGTATGAA ATTTGATTAA  
 9241 GTGATAAATT AGATTTCGAG CGTATGGCAG ATGTCAATTA TAGTGAACAA ATTA AAAACA  
 50 9301 TCGAAGTGAA TTTGAAATAA TCAATGTAT ATATAGAAAG AAAGCTTAAG AAGCGGTTTA  
 9361 ATAAATCCCAT GTTTAATGAT TTTGATACGT GTTTTTATAA TAAAAACATA TCGAACATTG  
 9421 ACTACGTTAT TAAGCTGCTT TTTTGTACAC TTTATAACCA ATAGCTTAAG ATTTAAAACT  
 9481 AATCGGAAAG AACAAATGAT CACCAAAAAA ATATTTATGT TGCTATTAAA AATCAGTTAA  
 55 9541 TACGAATGTT AAAATACGTT TGATTTTCAT TAATAATGAT TCAAGTTTTAT TTAATGAGC  
 9601 GTTAATGTCA GTCTGTTTTG ATGCACCTTA TAATAAAGAC AGATAGTTCA AATTACGTAA  
 9661 TAATAACAAT CCAATACATT AAGATTGGAG CAAAAAATA TGAAATTAAC AACGATAGCT  
 9721 AAAGCAACAT TAGCATTAGG AATATTAAT ACAGGTGTGT TTACAGCAGA AAGTCAAAC  
 9781 GGTCAACGCA AAGTAGAATC TGATGAGACA CAACGCAAT ATTTATATCA TATGCTACAT  
 9841 CAATACTATT CTGAAGAAAG TTTTGAACCA ACAAACATTA GTGTTAAAAG TGAAGATTAC  
 60 9901 TATGGCTCTA ACGTTTTAAA CTTTAAACAA CGAAATAAAG CTTTTAAAGT ATTTTFACTT  
 9961 GGTGACGATA AAAATAAATA TAAAGAAAAA ACACATGGCC TTGATGTCTT TGCAGTACCT  
 10021 GAATTAATAG ATATAAAGG TGGCATATAT AGCGTTGGCG GTATAACAAA GAAAAATGTG  
 10081 AGATCAGTGT TTGGATTGTT AAGTAATCCA AGTCTACAG TTA AAAAAGT TGATGCTAAA  
 65 10141 AATGGCTTTT CGATAAACGA GTTGTTTTTT ATTCAAAAGG AAGAAGTATC ATTGAAGGAA  
 10201 CTGGACTTTA AAATAAGAAA ACTCTTAATC GAAAAATATA GATTGTATAA AGGACGCTCT  
 10261 GATAAAGGTA GAATTGTTAT CAATATGAAA GACGAAAAGA AGCATGAAAT TGATTTAAGT  
 10321 GAAAAATTA GTTTTGAACG GTTTTGTGAT GTAAATGGAA GTAAAGCAAAT TAAAAATATT  
 10381 GAAGTGAATT TGAATTAGTT CGAGTTAATA GCATAATAGC TTAAGAAAGCG GCTTAACGAC  
 10441 AAAATGTGAA TTGGCATTTG TGTCTTTATA TAAGGAACGT TGTTAAATAC ATTACTGTTG  
 70 10501 TTAAGTTGTT TTTGTAATTC AAGAGCAGA ACAGAGTAAC ATCATCAGTT GTAGTAAACG  
 10561 ATAATCCGGT AAAACAACATA AATGAAATAA TGAAGTCAAT TTAACCTGAA CATTTAAATA  
 10621 TATTTGTTTT TCATTAAGAA TAATTCAAGT ATATTTAAAT CGAGGTTAAT TATCGTATGA  
 10681 AACGATGCAC GTTATAATAA AAATGTATGA TTCAAATTAC GTAAATGAAA CAATCCAATA

10741	TATTAAGATT	GGAGCAAATA	AATATGAAAT	TTACAGCATT	AGCAAAAGCG	ACATTAGCTT
10801	TAGGAATTTT	AACAACAGGA	ACTTTAACAA	CAGAAGTTCA	TTCAGGTCAT	GCAAAACAAA
10861	ATCAAAAGTC	AGTAAATAAA	CATGACAAGG	AAGCATTATA	CCGATACTAC	ACTGGAAAGA
10921	CTATGGAAAT	GAAAAATATT	AGTGCCTTGA	AACATGGTAA	AAACAACCTA	CGTTTAAAGT
10981	TTAGAGGTAT	TAAGATTCAA	GTTTTACTGC	CTGGAAATGA	TAAAAGTAAA	TTTCAACAGC
11041	GTA GTTATGA	GGGGTTAGAT	GTTTTCTTTG	TTCAAGAAAA	AAGAGATAAG	CACGATATAT
11101	TTTATACTGT	TGGTGGTGTA	ATACAGAATA	ATAAACATC	TGGAGTTGTC	AGTGCACCAA
11161	TATTAATAT	TTCAAAAGAA	AAGGGTGAAG	ATGCTTTTGT	GAAAGGTTAC	CCTTATTACA
11221	TTAAAAAGA	AAAAATAACA	CTAAAAGAAC	TGGATTATAA	GTTGAGAAAG	CATCTAATTG
11281	AAAAATACGG	ACTTTATAAA	ACAATCTCAA	AAGATGGTAG	GGTCAAAATT	AGCTTGAAAG
11341	ATGGCAGTTT	TTATAACCTT	GATTTAAGAT	CTAAATTAAA	ATTTAAATAT	ATGGGGGAAG
11401	TCATAGAAAG	CAAAACAAAT	AAAGATATTG	AAGTTAACTT	AAAGTAAATC	ATTACGAATA
11461	ATTAAGAGTA	ATTGAAGCGG	CTTAACGGTG	AAATGTAAAT	TGGTGCAGAT	AGCTTATACA
11521	AAAAGGATGC	ATCAATCGAT	ATCGTCGTTA	AGCCGTTTTG	GTTTGTGTGT	CATGAATCCT
11581	ATCCCAATGT	CCATAAAGGT	AAAATTTCCA	CCACCAACAT	CAAAATTCTC	CACATCCCAA
11641	CATAACCAAA	TGTTATAATA	AATCTATTAC	ACAAAGAGAT	AAATTACTTA	TTCAAAGGCG
11701	GAGGAATCAG	ATGTCATTAT	CTGAAAAACA	ACGTCAGCAA	CAAGCTGAAT	TACATAAAAA
11761	ATTATGGTTC	ATTGCGAATG	ATTTAGAGAG	GAATATGGAT	CGGAGTGAAT	TCCGTAATTA
11821	CATTTTAGGC	TTGATTTTCT	ATCGCTTCTT	ATCTGAAAAA	CGGGAACAAG	AAATATGCAG
11881	TGCCTTGTCA	GGTGAAGACA	TCACGTATCA	AGAAGCATGG	CGAGACGAAG	AAATACCGTGA
11941	AGACTTAAAA	GCAGAATTAA	TTGACCAAGT	CGGTTACTTC	ATTGAGCCAG	AAAGATTTAT
12001	CAGTCGGATG	ATTCGTGAAA	TTGAAACGCA	AGATTTTCGAT	ATCGAACACC	TGGCGACGGC
12061	AATTCGTAAA	GTTGAAACAT	CAACATTAGG	TGAAGAAAGT	GAAAATGACT	TTATCGGTCT
12121	GTTACAGCGAT	ATGGATTTGA	GTTCAACGCG	ACTAGGTAAC	AATGTCAAAG	AACGTACTGC
12181	TTTAATCTCT	AAAGTCATGG	TTAATCTTGA	CGACTTACCA	TTCGTTTACA	GTGACATGGA
12241	AAATGATATG	TTAGGTGATG	CATATGAATT	CCTAATTGGG	CGCTTTGCGG	CGACAGCGGG
12301	TAAAAAAGCA	GGCGAGTTCT	ATACACCACA	ACAAGTATCT	AAGATACTGG	CGAAGATTGT
12361	CACAGACGGT	AAAGATAAAT	TACGTCACGT	GTATGACCCA	ACATGTGGTT	CAGGTTCACT
12421	GTTGTTACGT	GTTGGTAAAG	AAACACAAGT	GTATCGTTAT	TTCGGTCAAG	AACGTACCAA
12481	TACTACATAC	AACTTAGCAC	GCATGAATAT	GTTATTACAT	GATGTGCGTT	ATGAGAACTT
12541	CGATATCCGT	AATGATGACA	CATTGGAAAA	CCCAGCCTTT	TTAGGCAATA	CATTTGATGC
12601	GTTTATTGCG	AACCCACCGT	ATAGTGCAGG	ATGGACTGCA	GATTCAAAGT	TTGAAAAATGA
12661	CGAACGATTG	AGTGGTTACG	GCAAACTTGC	GCCTAAGTCT	AAAGCAGACT	TTGCCTTTAT
12721	TCAACACATG	GTACATTACC	TAGACGATGA	AGGTACCATG	GCCGTTGTAC	TCCCACATGG
12781	TGTATTATTC	CGAGGTGCTG	CAGAAGGTGT	CATTCTGCTG	TATTTAATTG	AAGAAAAAGAA
12841	CTACTTAGAA	GCTGTGATTG	GTTTGCCAGC	GAATATTTTC	TATGGGACAA	GTATCCCAAC
12901	ATGTATTTTA	GTATTTAAAA	AATGTCGCCA	ACAAGACGAC	AACGTACTAT	TTATCGATGC
12961	ATCCAATGAT	TTTGA AAAAG	GAAAAAATCA	AAATCAATTA	AGCGATGCCC	AAGTCGAACG
13021	TATTATAGAC	ACATATAAGC	GTAAGGAAAC	AAATTGATAA	TATAGCTACA	GCGCGACACT
13081	ACAAGAGATT	GCCGATAACG	ATTACAACCT	AAATATACCG	AGATATGTCT	ATACATTCTGA
13141	AGAAGAAGCA	CCGATTGATT	TAGATCAAGT	CCAACAAGAT	TTGAAAAATA	TCGATAAAGA
13201	AATCGCAGAA	ATTGAGCAAG	AAATCAATGC	ATACCTGAAA	GAACCTGGGG	TGTTGAAAGA
13261	TGAGTAATAC	ACAAAAGAAA	AATGTGCCAG	AATTGAGGTT	CCCAGGGTTT	GAAGGCCAAT
13321	GGGAAGAGAA	GCAGTTAGGG	GATCTTACAG	ATAGAGTAAT	TAGGAAAAAT	AAAAACTTAG
13381	AAATCGAAAA	GCCTTTAACA	ATATCCGGAC	AGTTAGGTTT	AATTGATCAA	ACAGAATATT
13441	TTAGTAAATC	AGTTTCGTCT	AAAAATCTAG	AAAATTATAC	ACTAATAAAG	AATGGAGAAT
13501	TCCGCTATAA	CAAAAGTTAT	TCTAATGGAT	ACCCATTAGG	GGCTATTAAA	AGATTAACTA
13561	GATATGATAG	TGGTGTATTG	TCCTCTTTGT	ATATTGTGTT	TTCTATTAAA	AGTGAATGTT
13621	CTAAAGACTT	CATGGAAGCA	TATTTTGATT	CGACACACTG	GTATAGAGAA	GTTTCTGGAA
13681	TTGCAGTTGA	GGGTGCAAG	AATCACGGAT	TATTAAATGT	TTCTGTGAAT	GATTTTTTTA
13741	CTATTCTAAT	TAAATATCCA	AGTTTAGAAG	AACAGCAAAA	AATAGGCAAG	TTCTTACGCA
13801	AACTCGACCG	ACAAATTGAA	TTAGAAGAAC	AAAAGCTTGA	ATTACTTCAA	CAACAGAAAA
13861	AAGGCTATAT	GCAGAAAAAT	TTCTCACAGG	AACTGCGATT	CAAAGATGAG	AATGGTGAAG
13921	ATTATCCAGA	TTGGGAAAAAT	AGCAAAATAG	AAAAATATTT	AAAAGAGAGA	AACGAACGTT
13981	CTGACAAAGG	GCAAAATGCTT	TCAGTAACTA	TAAATAGTGG	CATTATAAAA	TTTAGTGAAT
14041	TGGATAGAAA	AGATAAATCA	AGTAAAGATA	AAAGTAATTA	TAAAGTAGTT	AGGAAAAATG
14101	ATATTGCATA	TAATTCATAT	AGAATGTGGC	AAGGGGCTAG	TGGTAAATCA	AATTATAATG
14161	GGATTGTTAG	CCCTGCATAT	ACTGTGCTTT	ATCCAACACA	AAATACTAGC	TCATTATTTA
14221	TTGGATATAA	GTTTAAAACA	CATAGAATGA	TTCATAAATT	TAAAATTAAT	TCACAAGGAT
14281	TACATTCAGA	TACATGGAAC	TTAAATATATA	AACAATTAAT	AAATATAAAT	ATAGATATAC
14341	CTGTATTGGA	GGAACAAGAA	AAGATAGGTG	ATTTCTTTAA	AAAAATGGAT	ATATTGATAA
14401	GTAACACAGAA	AATGAAAAAT	GAAATATTAG	AAAAAGAGAA	ACAAATCCTT	TTACAAAAAA
14461	TGTTCTTATA	ACTTTGATAA	ATACATAGAT	TGCATAAGAA	TAAAATTTGT	ATAATTTAAG
14521	ATAAAAGTTG	TAAAAGTAAA	GTGAATTAAA	AACGAACATT	AAATTTAGGC	ACTGTGAAAG
14581	CGCAGTGTCT	TTTTTGTGCT	GAAATTTGTT	ACAGAATAAG	TAGTTAAATA	AAGATTAAAG
14641	TGAGATAAAG	TGTTATTCTG	AAATAAAAGA	GAGTAGATCG	ATAGGAATTG	AATGATATTA
14701	GTTAACTATT	TATTAAATTA	CTTAATAATG	ATTAATTTTT	AGTTAAAGTA	AGTTTAAATG
14761	GAAGCACGAC	CATTGCTCAT	TATAATGAAT	GAGGATTGTT	CGTATTGCGT	AATAGAAATA
14821	ATCAAAATAGA	CTAAAAATCA	GGAGCATAGA	ATTATGAAT	TAAAAAATAT	TGCTAAAGCA
14881	AGTTTAGCAC	TAGGGATTTT	AACAAACAGG	ATGATTACAA	CTACTGCTCA	GCCAGTAAAA
14941	GCAAGTACAT	TAGAGGTTAG	ATCACAAGCT	ACTCAAGACT	TGAGTGAATA	TTATTAATAGA
15001	CCGTTCTTTG	AGTATACAAA	TCAGTCAGGA	TATAAAGAGG	AAGGAAAAAG	GACGTTTACT
15061	CCTAATTATC	AACTTATAGA	TGTAACTTTA	ACTGGGAATG	AAAAGCAAAA	TTTTGGTGAA

15121 GATATTTCTA ATGTAGATAT ATTTGTTGTA AGAGAAAATT CTGATAGATC TGGTAATACA  
 15181 GCTTCAATTG GTGGTATTAC TAAAACAAAC GGTTCAAATT ATATTGATAA AGTAAAGAT  
 15241 GTAAATTTAA TAATTACTAA AAACATCGAT AGTGTTACAT CAACGTCAAC ATCATCTACA  
 5 15301 TATACAATTA ATAAAGAAGA AATTTTCTTA AAAGAACTTG ATTTTAAATT AAGAAAGCAT  
 15361 TTAATTGATA AACATAACCT TTATAAGACA GAACCTAAAG ACAGTAAAT TCGAATTACT  
 15421 ATGAAAGATG GTGGGTCTTA CACATTTGAA TTGAATAAAA AGTTACAAAC ACACCGTATG  
 15481 GGTGATGTTA TTGATGGCAG AAATATAGAA AAAATTGAAG TGAATTTATA AAATTATTCTG  
 15541 AGGGAGCATA TCATGAGGGA AAATTTTAAG TTACGTAAAA TGAAAGTCGG TTTAGTATCT  
 10 15601 GTTGCAATTA CAATGTTATA TATTATGACA AACGGACAAG CAGAAGCATC TGAAATCAA  
 15661 AACGCTTTAA TCTCTAATAT AAATGTAGAC AATCAG

SEQ ID No. 105

*S. aureus* strain EMRSA 16(252) (SSL1-SSL11) -coding strand

15  
 1 TGAATGGCT TGATTTGTAA TATGTGTGCC TAAATGACCT GTAGCACCTG TTAACATAAT  
 61 ATTCATTAC TTCATCTCCT AATCTTTTATA TACATAACAT AATACTTAT CGATGATTTT  
 20 121 CAAAACATTT GATTTTATAA AAAATTGCAA TCTGTATTTA TTGTCGCCGT GTATAGTAAA  
 181 TACGCATATG TTATTAAATGT TAAAAATACC GTAATGACGC GTTTTAGTTG ATGCGTGCA  
 241 CCATGATCTT TGAAAAATTT GACATGGTAC TCGCAGCATA TGATGTCATT TTTGTGTCTG  
 301 AAAGTTTAC AGTTTTTAAA ATTAAAAATGG TTAAAGTGT GAATTGTATA AAAAGAGCT  
 361 TTGACGGATA AGAATTGATT ATTAACAGTT AGCATTTTAT TAATTACCTT AACAAATGATT  
 421 CAAGTTTAGT TAAATGAGGT TTAATTTGAA AGGGAATAGC GCCTCAATAT AATGTAGGTA  
 25 481 GATTGTTCAT ATTACGTAAT TGAAAAATCA AATTTAAATA GATTGGGGCT AAAAATTATG  
 541 AAATTTAAAG CGATAGCAAA AGCAAGTTTA GCATTGGGAA TGTAGCAAC AGGTGTAATA  
 601 ACGTCAATG TACAATCAGT ACAAGCGAAA ACAGAAAGTA AACAAACAAAG TGAGGCTGAT  
 661 TTAAGCTTT ATTATAATGG ACCAAGTTT GAATATAAAA AAGTAAGTGG ATATGGATTT  
 721 ATTGAAGGTA AAGATAGATT CATTGATTT ATATACAATG GACAATATAA TAAATATCT  
 30 781 TTAGTTGGTT CTGATAAAGA TAAATTAAT GAAGAAGTA ACCAGATAT AGATGTGTTT  
 841 GTCGTTAGAG AAGGAAACGG TAGACAAGCT GATAATCATT CGATTGGTGG CGTAACAAAA  
 901 ACTAATAGAG GAGTGTATTA CGACTATATA CACTCTCAA TCCTGAAAT TAAGAAAGGT  
 961 AATGAAGAAC CACAAAATAG TCTGTATCAA ATTTATAAAG AAAATATCTC ATTTAAAGAA  
 1021 CTTGATTATA GATTACGAGA ACGTGCAATC AAACAACACG GATTGTATTC AATGGCTCTT  
 35 1081 AAACAAGGTC AAATTACAAT TACTATGAAA GATGGCAAAT CACATACTAT CGATTTAAGT  
 1141 CAAAACCTTG AAAAAGAACG TATGGGTGAT TCTATCGACG GCAGACAAAT ACAAAAAATT  
 1201 CTAGTAGAAA TGAAATAATA CTTTCTAACG ACAAGCGCT ATGTTGAATC GTGCTTGTTA  
 1261 TGGAATATA TGGAAGTTAA GCGACGTACT GTTGCTTAGC TTCTTTTTTG AGGGCAAAGT  
 40 1321 TACAAAACCT ACACAAACAG TCGCAGCAGC CATTATCTTT TACTTAACTA GCTTAATCAT  
 1381 ATTTTATGAA TAGTTAAAAA AGGGTTAATG TGAATATCAG CATACAGCTC CTATAATATG  
 1441 GTTGATGAT TCAATTTACG TAATAAACA ATCTAATCTA ATATATTGGA GCATACAATT  
 1501 ATGAAAATGA AATCAATTGC AAAAGTAAGT TTAGTGTGG GTATTTTAGC TACAGGTGTA  
 1561 AATACTGTAA CGGAACAACC GGTGCATGCC GAAAATAAAC ATGTTCAAGT AAGCCAAAAT  
 45 1621 AGTAAAAATT TAAAAGCGTA CTATACTCAA CCTAGTGTG AATATAAAAA TGTGACAGGT  
 1681 TATATCAGTA GTATTCAACC TAAACCAGGC ACTAAATTTT TGAATATGAT AGAAGGTAAT  
 1741 ACAGTTAATA ATCTAGCTTT AGTTGGCAA GATAAGGAAC ACTATCATA GGGTGTACAT  
 1801 CGCAACCTTG ATATATTTTA TGTTAATGAG GATAAGAGAT TTGAAGGTGC AAAGTACTCG  
 1861 ATTTGTTGGTA TCACAAAAGC AAGCGATAAA GTTGTCGACC AAGTAGCAGA AGCAAGATC  
 50 1921 ATTAAGAAG ATCACAAGT TGAATATGAT TATGACTTTT TCCCATTAA AATAGATAAA  
 1981 GAAGCAATGA CATTAAAAGA GGTGATTTT AAAATAAGAA AACATCTTAT TGATCATTAT  
 2041 GGTCTTTATG GTGAAATGAG TTCAGGAACA GTTACTGTCC AAATGAAATA CTATGGCAAG  
 2101 TATACAATTG AATTGGATAA AAAGTTACAA GAAGACCGAA TGGCCGATAT CGTCAAAGTC  
 2161 ATAGATATTG ATAGGATTGA AGTCAAAGT AAAAAAGCAT AATGCTTAGA CTGGTCGTCG  
 2221 TAATGAATTG AAATTGCAAT AGTGAGGTTA AGTGATGATA AAACGTTACT TAACCTCTTT  
 55 2281 TTTATGTCTT AAAATCATTT CAAAGTACA TAGTAACACC ACATAAACCT CGAAGTACT  
 2341 CATTATTTTT TGCTTAAATG ACTTAATAAT ATTTCAACAA TTGTTAAACG AGGGTTAATG  
 2401 TGAATAGACC AATATGTCCT CTATAATGAA GTTGATGAT TCTAATTACG TAAAGAACAA  
 2461 TCGAATAATT ACGATTGGAG CATACAATA TGAAAAAAG AACAAATTGCG AAAGCGAGTA  
 2521 TAGCGTTAGG GCTTGTAAAC ACAGGTGCAA GTATAGTAAC AACGCAATCG GCCAATGCAG  
 60 2581 AAGTAGCATC AGCACTTAAA GCAGGACAGT TAGCAAAGAT AAACGTATCA ACAAGCGCAA  
 2641 TCACAACGAC AGCGCAAGCA GTGAACGACG AACAAAATCA TACGGCTAAT CCTGAACAGG  
 2701 CGGCAAAAGT TAACACAGAA AATGTATCAA CATCGCTTTC AACTCAAACA GAACAGACGC  
 2761 CAAAATCTAA CATACAAAAA GCAACCAAC CAGCACCAAG CGGAACAAC CTGCTTTTAC  
 65 2821 CAAATGCACA ACCGCAATCA ACACAACCAA CACCAAGTGT AACAAACCCG CCTTCATCTA  
 2881 ATGTCGAAAC ACCACAACCA ACATCGCCAA CCACAAAAAG AGCACAAGAG GAAATAAAC  
 2941 CTAAATTTAA AGGTTTAAGA TCATATTATA CGAAATCAAG TTTAGAATTT AAAAATGAGT  
 3001 TGGGTATAAT TATCAAAAAA TGGACAACAA TAAGATTTAT GAATATTGTT CCAGATTATT  
 3061 TCAATATATA AATAGCTTTA GTTGGAAAAG ATGATAAGAA ATATGGTGAA GGCGTGCATA  
 70 3121 GGCATGTGCA TGTATTTATC GTTTTGAAC AAAATAAATA TGGCGTAGAC AAATACTCGG  
 3181 TCGGTGGTAT CACAAAAGCA AATGAAAAGA AAGTTGATTA CAAAATGGA ATAAGTATTA  
 3241 CTAAAGAAGA TAAAAAGGT ACAATCTCAC ATGATGTTTC AGAATACAAG ATTACTAAAG  
 3301 AAGAGATTTC CTTGAAAGAA CTTGATTTTA AATTGAGAAA ACAACTCATT GAACAACATA  
 3361 ATTTGTACGG TAATATTGGT TCAGGAACAA TCGTTATTAA AATGAAAAAT GGTGGAAGT

3421 ACACGTTTGA ATTACACAAA AAATTGCAAC AACATCGCAT GGCAGATGTT ATTGATGGTA  
 3481 CAAATATTGA TAGAATTGAA GTGAATCTAA AATCATCGTG ATGGATACTA CATAGAACCG  
 3541 TTGAACGATG AAATGAGAAG TTAAGCGACA ATGACCTACA TCATTGCTTA GCTTCTTTTG  
 3601 TTTTGCCTAA TGATGTTAAG GGAAATTATT CACTTGTTTG TAAAAGTGAC ATCTCAATGT  
 5 3661 CTTGAAAATG ATGCAATGCC CAATGTGCAA GCGTTGAATC GCTTCGAAAC CATCTGTATT  
 3721 TAACGAAAAAT TATGGATTAC TTAACGATGA TTCAAATATA GTTAAACATG GTTTAATGTG  
 3781 AATAGACCGA TACGTCTTCT ATAATGTAGT TGTATGATT CTAATTACGTA AAGAACAATC  
 3841 GAATATAATA GATTGGAGCA TACAACATG AAAATGACAC AAATTGCGAA AACCAGTTTA  
 10 3901 GCACTATCAC TTTTAACAAC AGGTGCAAGT ATAGTAACAA CGCAATCGGC TAATGCAGAA  
 3961 GTAGCATCAG CACTTAAAC AGAACAGGCA ATAAATCTA AAATACAAA AGTAACAACA  
 4021 TCGCCTTCAC CAAATGTACA ACCGCAATCG CCACAACCAA CACCAAGTGT AACAACTCCG  
 4081 CCTTCACCAA ATGTACAACC GCAATCGCCA CAACCAACAC CAAACCCAAC AACACCGCAT  
 4141 TCATCTAATG TCGAAACAAA ACAACCGCAA TCGCCAACCA CAAAACAAGC ACAAAGGAA  
 15 4201 ATAAACCCCTA AATATAAGGA TTTAAGAACA TATTATACGA AACCGAGTTT AGAATTCGAA  
 4261 TATTCGTTGCG GGGGTATCAC GAAGACAAAT AGTAAAAAAG TTGATCACA AGCAGAATTA  
 4321 GATTGGTTCA TTTATAAAAT AGCTTTAGTT GGAAAAGATG ATAAGAAATA TAAAGATGGG  
 4381 CCATACGATA ATATTGATGT ATTTATCGTT TTAGAGGACA ATAAATATCA ATTGAAAAA  
 4441 TATCTGTGTCG GGGGTATCAC GAAGACAAAT AGTAAAAAAG TTGATCACA AGCAGAATTA  
 20 4501 AGCATTACTA AAAAAGATGA GAAAGGCAAA ATCTCACACG ATGATTCAGA ATATAAGATA  
 4561 ACTAAAGAAG AGATTTCCCT GAAAGAGCTT GATTTTAAAT TGAGAAAGCA ACTCATTGAA  
 4621 CAACATAATT TGTACGGTAA TATTGGTTCA GGAACAATTG TTATTAAAC GAAAACCGGT  
 4681 GGAAGATACA CGTTTGAATT TCATAAGAAC ATCGCATGGC AGATGTCATA  
 4741 GATGGCACA GTATCGAAAG AATTGAAGTG AATCTAAAT CATCGTGATG GATACTGCAT  
 25 4801 AGAACCGTTG AACGATGTAC CGAGAAGTTG AGTGACAAAA ATTTACATGT TGCTCAGCTT  
 4861 CTTTCGTGTT GCACAGTGGT ATAAGAAGAC GAATATTTGA TGTGACCAC ACATAACATG  
 4921 CTTATGTCCT AAATGTGTCT AAAAATACCA ATCTATCAAG CGTTGAAGTA CAACAAAAGT  
 4981 CATTTTTAT TAAACGAACAT TATGGATTCC TTAATTTAAT TAACGATGAT TCAAAATAG  
 5041 TTAACAAGG TTTAATATGA ATAGAGACAT ACGCCAATA TAATAAAGCT GTATGATTCA  
 30 5101 ATAAACGTAA TCGAACAAAT CTAATAATTA CGATTGGAGC ATACAACATAT GAAGATGACA  
 5161 GCAATTGCGA AAGCCAGTTT AGCTCTAAGT ATTTTAGCGA CTGGGGTTAT AACATCAACG  
 5221 GCTCAAACTG TAAATGCGAG CGAACATGAA TCAAAATATG AAAATGTGAC AAAAGATATC  
 5281 TTCGATTGGA GAGATTACTA TAGTCGCGCA AGTAAGGAAC TTAATAATGT TACTGGCTAT  
 5341 CTTTATAGCA AAGGTGGCAA GCACTACCTT ATCTTTGATA AAAATAGAAA GTTCACAAGA  
 35 5401 ATACAAATTT TTGGTAAAGA TATAGAAAGG ATTAATAAAC GTAAAAATCC AGGATTAGAT  
 5461 ATATTTGTAG TTAAGAAGC GGAATAATCG AACGGCACAG TGTATTCTATA TGGTGGCGTT  
 5521 ACTAAGAAAA ATCAAGGCGC TTACTATGAT TACTTAAGCG CACCTAGATT CGTTATTAAA  
 5581 AAAGAAGTAG GTGCAGGTGT TTCTGTGCAT GTTAAAGAT ACTACATTTA TAAGGAAGAG  
 5641 ATTTGCGCTGA AAGAACTTGA TTTTAAATTA CGACAGTATT TAATTCAGA TTTTGTCTGT  
 40 5701 TATAAAAGT TTCCTAAAGC TAGCAAGATA AAAGTGACAA TGAAGATGG CGGCTATTAT  
 5761 ACGTTTGAAC TCAATAAAAA GTTACAAACA AACCAGATGA GTGACGTCAT AGATGGCAGA  
 5821 AATATTGAGA AAATAGAAGC CAATATTAGA TAATTCACAG AACATGGAT AATAGTAAAA  
 5881 TATGGATAGT ATAGATGAGC TAGGCAGCAT AGGTTGCTTA GCTTCTTTTT TGTGTCGTTA  
 5941 CGATCGGAAA TGAAGTGGGT CGATGAAAGA TAATAACGAC ATAAAAAATG GTGAAAATAG  
 45 6001 TTGATACCTA TAGTCGCACG GTGTGCGACT AGTGACATGA AACAAATGTGG AAAACATAAT  
 6061 TAAATTGAAA GAAAGTGTGA ATGGTGAAAA AAATAGCATA GAGTTATAAA AAATGATTAA  
 6121 TACTGTTAGG ATTCCATTAA TTAGCTTAAC ATTTGGTTCAG GAATAGTTAA AAAGCGGTTA  
 6181 ATTCATAGCG CAGTATCCGG CTTATATAAT GATAGTAGAT TGTTCTGATT ACGTAAATTGA  
 6241 ATTAATCACA TAAACATATA TTAAGTCAAA ATTTAAAAAT AGATTGGGAG AATAATACTA  
 6301 TGAAATTTAA AACGTTAGCT AAAGCAACAT TAGCATTAGG TTTATTAAC ACTGGTGTCA  
 50 6361 TTACATCAGA AGGTCAAGCA GTTCAAGCGG CAGAAAAACA AGAGAGAGTA CAACATTTAC  
 6421 ATGATATTAG AGATTTACAT CGATACTACT CATCAGAAAG TTTCGAATAT AGTAATGTTA  
 6481 GTGGTAAGGT TGAAAACCTAC AATGGTTCTA ACGTTGTACG CTTTAACCCA AAAGATCAAA  
 6541 ATCACCAAAT ATTCTTATTA GGAAGAATA AAGAACAATA TAAAGAAGGT CTACAAGGCC  
 55 6601 AAAATGTCTT TGTAGTACAA GAATTAATTG ATCCAAACGG CAGACTATCT ACTGTTGGTG  
 6661 GTGTAACGAA GAAAAACAAC AAACTTCTG AAACATAAC ACCTTTATTT GTTAATAAAG  
 6721 TTAATGGTGA AGATTTAGAT GCATCAATTG ACTCATTTTT AATCCAAAAA GAAGAAATCT  
 6781 CATTAAGAAG GCTTGATTTC AAAATTAGAC AACAATTAGT TAATAATTAC GATTATATA  
 6841 AAGGTACATC TAAATACGGT AAAATCATTA TCAATTTGAA AGACGAAAAA AAAGTAGAAA  
 6901 TTGATTTAGG TGATAAATTA CAATTCGAGC GCATGGGCGA TGTGTTGAAT AGTAAAGACA  
 60 6961 TTAGAGGTAT ATCAGTCACT ATTAACCAAA TTTAAAGTAA GCAATCAATG ACTTTAAAGT  
 7021 AATAAATTTT GAGGCAGCTT AACGATGAAA CGTTGAATAA ATATGTATAT CCTATCAAG  
 7081 GAGCGTATTT AAAGCAACGT CGTCGTTAGG CTGTTTTTTA CATTTGATAA CACAAGTCAC  
 7141 GAGTATACCTA AATAAGTCAT ACCACTCATT AAAACACAGT ACTTTGAATA TTAATTAGGT  
 7201 GTTCATTAA AATGATTCAA AGTTATTTAA ACGCACGTTA ATGTCAGTTT GTTTCGATGC  
 65 7261 ACTTTATTAAT AAAGGCAGAT GTTTCAAATT ACGTAATCAT AACATCCAA TACATTAAGA  
 7321 TTGGAGCAAT AAATTATGAA ATTAACAGCA ATAGCTAAAG CAACATTAGC ATTAGGAATA  
 7381 TTAACAACAG GTGTGATGAC AGCAGAAAGT CAACTGTAA ACGCGAAAGT AAAGTTGGAT  
 7441 GAAACACACG GCAAAATATTA TATAAATATG CTAAGAAGAT ACTATCTCA AGAAAGCTAT  
 7501 GAATCAACAA ACATTAGTGT TAAAAGTGAA GATTATTATG GGTCTAACGT TTTAACTTT  
 7561 AACCAACGAA ATAAAAATTT CAAAGTATTT CTAATTGGCG ACGATAGAAA TAAATATAAA  
 7621 GAAGTACGCG ATGGCCGTGA TGTCTTTGCA GTACCTGAAT TAATAGATAC TAAAGGCGCG  
 7681 ATATATACCG TTGGCGGTAT AACAAAGAAA AATGTGAGAT CCGTGTGTTG CTATGTAAGT  
 7741 CATCCGGGAC TGCAAGTTAA AAAAGTTGAT CCTAAAGATG GCTTTTCGAT AAAAGAGTTG

7801	TTCTTTTATTC	AAAAAGAAGA	AGTATCATTG	AAAGAACTTG	ATTTTAAAT	CAGAAAAATG
7861	TTAGTCGAAA	AATATAGATT	GTATAAAGGC	GCGTCAGATA	AAGGTAGAAT	TGTTATCAAT
7921	ATGAAAGACG	AAAAAGAAACA	TGAAATTGAT	TTAAGTGAAA	AAATTAAGTTT	TGATCGTATG
7981	TTTGTATGTC	TGGATAGTAA	GCAGATTAAA	AAATATTGAAG	TGAATTTGAA	TTAATTTAAAG
8041	ATGATAGTAT	AAAAAATTAA	AAAGCGGCTT	AACGATAAAA	TGTGCATTGA	CACCCGTACC
8101	TTTGAATAAG	AACTGTGTTA	AAATACATTAT	TGTTGTAAAG	TTGTTTTTTG	CGTTTTCAAAG
8161	AGTAGAATA	ACACAATTAG	TTGTAGAAAA	CGATGATCCG	GAAAAACAAC	GACATGTAAA
8221	GCGTGAAAGT	TAATTAACCT	GGACATTAAA	ATATATTTGT	TTTTAATTAA	TAATAATTCA
8281	CGCATATTTA	AATCGAGGTT	AATTATCGCG	TTAAACGATG	GACGTTATAA	TAAGCGTATA
8341	TGATTCAAAT	TACGTAATAA	TAACAATCCA	ATACATTAAAG	ATTGGAGTAA	ATAAATATGA
8401	AATTGACAGC	ATTAGCAAAA	GTAACATTAG	CATTAGGGAT	TTTAACAACA	GGAACTTTAA
8461	CAACAGAGGC	CCATTCAGGT	CATGCTAAAC	AAAATCAAAA	GTCAGTAAAC	AAACATGACA
8521	AAGAAGCATT	ACACCGATAC	TACACTGGAA	ACTTTAAGGA	AATGAAAAAC	ATTAATGCTT
8581	TGAGACATGG	TAAAAATAAC	TTACGTTTTA	AAATATAGAGG	GATGAAGACT	CAAGTATTAT
8641	TGCCGTGGAA	TGAGTACCGT	AAATATCAAC	AGCGAAGACA	TACGGGCTTA	GATGTGTTTT
8701	TTGTTCAAGA	AAGAAGAGAC	AAGCAGGACA	TATCATATAC	TGTTGGTGGT	GTAAACAAGA
8761	CCAATAAAAC	ATCCGGCTTT	GTCAGCACAC	CAAGATTAAA	TGTTACAAAA	GAAAAGGGTG
8821	AAGATGCTTT	TGTGAAAGGT	TACCCCTATG	ATATTAATAA	AGAAGAAATA	TCATTGAAAG
8881	AGTTAGATTT	TAAGTTGAGA	AAGCATCTAA	TTGAAAAATA	CGGTCCTTAT	AAAACACTCT
8941	CAAAAGATGG	TAGGATTAAA	ATTAGTCTGA	AAGATGGTAG	CTTTTATAAC	CTTGATTATA
9001	GAACATAAAT	AAAATTCAAA	CATATGGGGG	AAGTCATAGA	TAGTAAACAA	ATTAAGGATA
9061	TTGAAGTGAA	TTTAAAGTAA	ATCATTATGA	ATAATAAAAA	GTAATTGAAG	CGGCTTAACG
9121	ATGAAAAGTA	AATTGGTGCG	TATACCTTAC	CAAAAGGATG	CATCAATCGA	TATCGTCGTT
9181	AAGCCGTTTT	TGTTTGCGTG	TTATGAATCC	TATCCCAATC	TCCATGAATA	TAAAAATTC
9241	ACCATCAACA	TCAAAATCT	CAACATCGCG	ACACCACCAA	ATGTTATAAT	AAATCTATTA
9301	CACAAAGAGA	CAAATTACTT	ATGCAAGGCG	GGAGGAATCA	CATGTCTATT	ACTGAAAAAC
9361	AACGTCAGCA	ACAAGCTGAA	TTACATAAAA	AATTATGGTC	AATTGCGAAT	GATTTAAGAG
9421	GGAATATGGA	TGCGAGTGAA	TTCCGTAAAT	ACATTTTAGG	CTTGATTTTC	TATCGCTTCT
9481	TATCTGAAAA	AGCGGAACAA	GAATATGCAG	ATGCGCTGGC	AGGTGAAGAT	ATTACATATC
9541	AAGAAGCATG	GGCAGATGAA	GAATACCGTG	AAGACTTAAA	AGCAGAATTA	ATTGATCAAG
9601	TTGGTTACTT	CATTGAGCCA	CAAGATTAT	TCAGTGCAGT	GATTTCGTGA	ATTGAAACGC
9661	AAGATTTTCA	TATCGAACAT	CTGGCGACGG	CAATTCGTAA	AGTTGAAACT	TCAACACTAG
9721	GTGAAGAAAG	TGAAAATGAC	TTTATCGGAC	TGTTACAGCGA	TATGGACTTA	AGTTCAACGC
9781	GTTTAGGTAA	CAATGTCAAA	GAACGTACTG	CGTTAATTTT	CAAAAGTTATG	GTTAACCTTG
9841	ATGACTTACC	ATTCGTTTAC	AGTGATATGG	AAATTGATAT	GTTAGGTGAT	GCATACGAAT
9901	TTCTTATCGG	GCGCTTTGCG	GCGACAGCGG	GTAAAAAGC	AGCGCAGTTC	TATACACCAC
9961	AACAGACTAT	TAAAGTACTG	GCGAAGATTG	TCACAGACGG	TAAAGATAAA	TTACGTCATG
10021	TATATGACCC	AACATGTGGT	TCAGGTTTAT	TACTGTTACG	CGTTGGTAAA	GAGGCAAAAG
10081	TGATATCGTTA	TTTTGGACAA	GAACGTAAAC	ATACCACATA	CAACTTAGCG	CGCATGAACA
10141	TGTTGTTTAC	TGATGTGCGA	TATGAAAATT	TCGATATCCG	TAATGATGAT	ACGTTGGAAA
10201	ATCCAGCCTT	TTTAGGACAT	ACATTTGATG	CGGTTATTGC	GAACCCACCA	TATAGCCGCA
10261	AATGGACAGC	AGACTCAAAA	TTTGAATATG	ACGAACGCTT	CAGCGGATAC	GGCAAGCTTG
10321	CGCCAAAGTC	TAAAGCAGAC	TTTGCCCTTA	TTCAACACAT	GGTACATTAC	CTAGATGATG
10381	AAGGTACCAT	GGCCGTTGTA	CTCCACATG	GTGCTTTATT	CCGTGGTGCC	GCAGAAGGTA
10441	TCATTGCTCG	TTATTTAATT	GAAGAAAAGA	ACTACTTAGA	AGCCGTGATT	GGTTTGCCAG
10501	CGAACATTTT	CTATGGGACA	AGTATTTCAA	CATGTATCTT	AGTATTTAAA	AAATGTCGCC
10561	AACAAGACGA	CAACGTATTA	TTTATCGATG	CATCCAATGA	TTTTGAAAAA	GGAAAAAACC
10621	AAAGACATT	AAGCGATGCC	CAAGTCGAAC	GAATTATAGA	CACATATAAG	CGTAAGGAAA
10681	CAATTGATAA	GTATAGCTAC	AGCGCGACAT	TACAAGAGAT	CGCCGATAAC	GATTACAACT
10741	TAAACATACC	GAGATATGTC	GATACATTCC	AAGAAGAAGC	GCCAATTGAT	TTAGATCAAG
10801	TCCAACAAGA	TTTGAATAAT	ATCGACAAAG	AAATCGCAGA	AATTGAACAA	GAATCAATAG
10861	CATACCTGAA	AGAACTTGGG	GTGTTGAAAG	ATGAGTAATA	CACAAACGAA	AAATGTGCCA
10921	GAGTTGAGAT	TCCCAGGGTT	TGAAGCGCAA	TGGGAAGAGA	AGAAGGTTGG	CGAGTTATTA
10981	GAATTTAAAA	ATGGTTTAAA	TAAAGGAAAA	GAATATTTTG	GCTCAGGATC	GTCGATTGTT
11041	AACTTCAAAG	ATGTATTTAA	TAACAGGAGC	TTAAATACAA	ATAATCTGAC	TGGAAGAGTT
11101	AATGTGAATA	GCAAAGAACT	AAAAAATTAT	TCTGTTGAAA	AGGGTGATGT	TTTTTTTACA
11161	AGGACTAGTG	AGGTAATTGG	TGAAATAGGT	TATCCGCTCT	TAATTTTAAA	TGACCCTGAA
11221	AATACTGTGT	TTAGTGGATT	TGTATTAAGA	GGGCGGCCCTA	AATCAGGAAT	TGATTTAATA
11281	AATAATAATT	TTAAAAAGATA	TGTCTTTTTT	ACTAATTCAT	TTAGAAAAGA	AATGATTACA
11341	AAAAGTTCTA	TGACAACTAG	AGCTTTAACA	TCAGGTAGCG	CAATTAATAA	AATGAAGGTC
11401	ATATACCTTG	TTTCGGCTAA	AGAACAGAGA	AAATAGGTTG	ACTTCTTCAG	CAAACTCGAC
11461	CGACAAATTG	AAATTAGAAGA	ACAAAAGCTT	GAATTACTTC	AACAACAAAA	AAAAGGCTAT
11521	ATGCAGAAAA	TCTTCTCACA	GGAAGTGGCA	TTCAAAGATG	AGAAATAGTG	AGATTATCCA
11581	CATTGGGAAA	ATAGCAAAAT	AGAAAAATAT	TTAAAAGAGA	GAAACGAACG	TTCTGACAAA
11641	GGTCAAAATG	TTTCAGTAAC	TATAAATAGT	GGCATTATAA	AAATTTAGTG	ATTGGATAGA
11701	AAAGATAATT	CAAGTAAAGA	TAAAAGTAAT	TATAAAGTAG	TTAGGAAAAA	TGATATTGCA
11761	TATAATTCTA	TGAGAATGTG	GCAAGGGGCT	AGTGGTAGAT	CAAATTATAA	TGGGATTGTT
11821	AGCCCTGCAT	ATACTGTGCT	TTATCCAACA	CAAAATACTA	GCTCATTATT	TATTGGATAT
11881	AAGTTTAAAA	CACATAGAA	GATTCATAAA	TTTAAATTTA	ATTCACAAGG	ATTAACATCA
11941	GATACATGGA	ACTTAAATA	TAAACAATTA	AAAAATATAA	ATATAGATAT	ACCTGTATTG
12001	GAGGAACAAG	AAAAGATAGG	TGATTTCTTT	AAAAAATG	ATATATTGAT	TAGTAAACAG
12061	AAAATAAAAA	TTGAAATATT	AGAAAAAGAG	AAACAATCCT	TTTTACAAAA	GATGTTCTTA
12121	TAACTTTGAT	AAACACATAG	GTTGCATAAG	AAATAAATTT	GTGTAATTTA	ACATAAAAGT

12181 TATAAACATA AAGTAAATTA AAAACGAACA TTAAATTTAG GCACTGTGAT AGCACAGTGT  
 12241 CTTTTTTGTG TCGAAATTGT GTACAGAATA AGTAGTTAAA TAAAGATTAA GTTGAGATAA  
 12301 AGTGTATTTC GTAAATAAAA GAGAGTAGAT CGATAGGAAT TGAATGATAT TAGTTAACTA  
 12361 TTTATTAAAT CACTTAATAA TGATTAATTT TTAGTTAAAG TAAGTTTAAAT GTGAAGCAGC  
 12421 ACCGTTGCTC ATTATAATGA ATGAGGATTG TTCGTATTGC GTAATAGAAAT AAATCAAATA  
 12481 GACTAAAAAT TGGGAGCATA GAATTATGAA ATTAAAAAAT ATTGCTAAAG CAAGTTTAGC  
 12541 ACTAGGGATT TTGACAACAG GGATGATTAC AACTACTGCT CAGCCAGTAA AAGCAAGTGA  
 12601 GCAAAGCAGA TTATCAGTTA CTTCAAACGA CACGCAAGAA TTAAAAAAT ACTACAGTGG  
 12661 AACAGGATAT AATTTTCAAA ATGTGAGTGG TTATAGAGAA AAGGATAAAA TGAACATTAT  
 12721 TGATGGGACA CAACCTAATG TAGTTACTTT ACTTGGTACA GACAAAGAAA GATTTAAAGA  
 12781 TTATGACTAT GATTATGAAG GGTTAGATGT CTTTGTAGTC AGAGAAGGAT CAGGTAAACA  
 12841 AGCTGAAAAAT ATTTCAATAG GTGGAATTAC CAAGACGAAT AAAACGATT ATAAAGATTT  
 12901 CGTAAATAAT GTAGGTCTAG AAATAACTAA ACCAACAGGA CATAATACAG CACAAGACA  
 12961 AGCAGAAACT TATAGAATTA ATAAAGAAGA AATTTTATTA AAAGAATTAG ATTTCAAATT  
 13021 AAGAAAACAT TTAATTGAAA ATCATGAAC TTTATAAGACA GAGCCTAAAG ACGGTAAAT  
 13081 TAGAATTACT ATGAAAGGTG GCGGCTACTA TACTTTTGAA TTAATAAAA AATTACAGCC  
 13141 TCATCGTATG GGTGATGTAA TTGATGGTAG AAATATAGAA AAAATTGAAG TCGATTTATA  
 13201 TTAATATTCG AGGGAGTATA TCATGAGGGA AAATTTTAAAG TTACGTAAAA TGAAGTCGG  
 13261 GTTAGTATCT GTTGCAATTA CAATGTTATA TATCATGACA AACGGACAAG CAGAAGCATC  
 13321 AGAGGCTAAT GAGAAGCCAA GTACAAATCA AGAATCAAAA GTTGTTCAC AGACTGAACA  
 13381 AAATTCAAAA GAAACAAAAA CAGTAGAATC TAATAAGAAC TTTGTTAAAT TAGATACTAT  
 13441 TAAACCTGGA GCTCAAAAGA TAACGGGAAC TACTTTACCA AATCACTATG TTTTATTAAAC  
 13501 AGTTGATGGG AAAAGTGCGG ATTCAGTAGA AAATGCGGGT TTGGGTTTGT TTGAAGCAAA  
 13561 TGACAAAGGA GAATTTGAGT ACCCTTTAAA TAATCGTAAA ATTGTTTATA ATCAAGAAAT  
 13621 TGAGGTTTCG TCGTCAAGCC GTGATTTAGG TGAAGATGAA GAAGATGAAG AGGTGGAAGA  
 13681 AGCTTCAACT GATAAAGCTG GC

SEQ ID No. 106

30 *S. aureus* strain EMRSA 16(252) (SSL12-SSL14)

1 CTACATTTAA AGTATAACTA TTTTCAAGT AGTTAGAAAA TTCAATAAAT AACTAATTTA  
 61 ACGCAATTAA TATCATAATT AGCATCAGTT TTAATTTAT AGCACTCTTT TCATTTTATA  
 121 ATTGGCAATG TTCCAATGCT TAAAAATGAA AAATTTTACT TAAATCACAC GTTCAACTGT  
 181 TTATGTTTAT TTAACATAGT GTTAGCTTTT ATTTAATTCC GAATCGGTGC AATACCTTAT  
 241 ATACTATGGG TAATTCACGC AAAGGAGATT ATCATATGAA AAAGAACATC ACGAACAAAA  
 301 TTGTTTTATC AACAGCATTG TTACTTTTAG GAACATACAT AACACAACCG CCTAAAGCAC  
 361 CAATCAGTTT TTCATCTGAA GCAAAAGCAT ACAATATAAG TGAAACGAA ACGAATATCA  
 421 ATGAATTAAT CAAATATTAC ACGCAGCCAC ACCTTTCATT TTCTGGAAAA TGGCTATGGC  
 481 AAAACCCATA CGGTACCATA CATGCAACAT TACAAACGTC GGTGTTGGTAC ACGCATATTC  
 541 AAGTATTTGG TCCTGAAAGT TGGGGCAACA TTAATCAATT AAGAGATAAA TATGTAGATA  
 601 TTTTTGTTAC CAAGGATGAG GATACTATAG AAGGCTATTG GACTTATGAT GAAACCTTTA  
 661 CTGGAGGTGT AACACCAGCA GCTACTTCAT CTGATAAGCC TTATAAATTA TTTTAAAAAT  
 721 ATAAGGATAA ACAACAACT ATGATTGGTG GACTTGAGTT TTATCAAGGT AATAAACCCAG  
 781 TAATAACTTT AAAAGAACTA GATTTCCGCG TACGTCAAAC TTTAATTAAG AACAAAAAGC  
 841 TATATAATGG AGAATTTAAT AAAGGTCAAA TTAGGATTAC AGGTGGTGGT AATCATTATA  
 901 CGATTGATTT GAGTAAAAAG TTAACAACTA CTGACACAAA CAGTTATGTT AAAATCCTC  
 961 GACATGCAGA AATTGAAGTC ATTCTTGAAA AATCTCATTG GTATCATTCA TTGAGTGAAT  
 1021 ACAAAATAGAC TGAATGATTT AATTGATTTT TCTCTTTTAA TAACTTATAT ATTCCTCGCA  
 1081 TATAACATTG CTCGTAAGAG AGATTTTCTA ATGAAAAACA ACCTCACAAA AAAAATTATAT  
 1141 TTATCAATAG CATTTGACTAT GATAGGCACA GCTACAAGTC AATCCCTCTA TTCACCGATT  
 1201 AGTATGTCAT CTGAAGCAAA AGCATACAAT ATAAGTGAAA GCGAAACGAA TGTAAATGAA  
 1261 TTGACTAAAT ATTACACACA ACGCCATTTA ACATTTTCAA ATAAATGGCT ATGGCAAAAA  
 1321 GATAATGGTA CAATCCATGC AACATTGTTA CAGTTATCAT GGTTTAGTCA TATTCAGTA  
 1381 TTCGGCCCTG AAAGTTGGGG TAATATTAAT CAATTAAGAA ATAAATATGT AGATATTTT  
 1441 GCTTTAAAAAG ATTATGAAAC CTGGCGCACT TATATGTTAG CCCAAGAAAC ATTTACTGGC  
 1501 GGTGTTACAC CAGTAGCAAA ACCTAGCGAT AAACATTATA AATTGAATGT AACATATAAA  
 1561 GATAAAGCAG GAACATTTAT TGGCGAATAT CAATTTTATA CAGGTAATTA ACCAGTTTAA  
 1621 ACTTTAAAGG AAGTTGATTT CCGAATCCGC CAAACACTAA TTAATAACAA AAAATTATAT  
 1681 AATGGTGATT ATAATAAAGG CCACATTAAG ATTACAGGTG GCAGCAACAA CTACACAATA  
 1741 GATTTAAGCA AAAGATTAAA ATCAACTGAT GCAAAATAGAT ATGTTAAAAA CCTTCAAACT  
 1801 GCACACATCG AAATCCTCCT TGAAAAATCT AGCTAACATC GAAATCGAAG TAAACAGAAA  
 1861 TAAGCACAAT GTTTACATTG ACGCCGTTC TTTTAAATTT TGTTAATTCCT CGTATTGATG  
 1921 ATTACGCAGA AAGGAGAAGT TCGAATGAGA AAGTGCATTA TGAAAAAGTT AATTTTAAATG  
 1981 ACAACATTGT TATTATTAGG AACAACAGCT ACACAACCT CTAATTCGCC ATTAATGTTT  
 2041 TCTACAGACG CTAAAGCTTA TCATATTGGC CAAGACGAAA CGAACATCAA TGAGTTAATT  
 2101 AAATATTACA CACAACCTGC TCTCACAATT TCAAAATAGAT GGTATATACCA ATATGATGAC  
 2161 GGCAACATTT ATGTTGAATT TAAGCGTTAT TCTTGGTCAG CGCATATACG ATTATGGGGT  
 2221 GCTGAAAGTT GGGGCAACGT TAATCAATTA AGAGACCGCT ATGTAGACGT GTTGGTTTAA  
 2281 AAAGACGAAG ACACTAGCCA GTTATGGTGG GTATATCGTG ATACATTAC AGGTGGTGT  
 2341 ACACCTGCTG CATCACCTTC TGATAAACCA TATAGTCTTT TTGTCCAATA CAAAGATAAA  
 2401 CTACAAACAA TTATTGGTGC ACATAAGATG TACCAGGTA ATAAACCAAT ATTAACATTG  
 2461 AAAGAAATTG ATTTTCGAGC ACGTGAAACA CTAATTAATA ATAAATATAT ATATCATGAA

5 2521 AATCGCAATA AAGGTAACT TAAATCACC GGTGGCGGCA ATGACTTTAC AATGATTTG  
 2581 AGTAAAAGAT TACATTCTGA TCTTGCAAT GTTTATGTTA AAAATCCTCA AAAAATAACA  
 2641 GTCGAAGTCC TCATTGATTA AATAAGGTCG TATGATTAT ATCGTAAATA CTAAGGATAA  
 2701 ACAATATTTG AGTAATACTA CAAAATGTCT CAATAGTACT TACACAATCA TTAATTTAGT  
 2761 ACACAGAATA TTAATGCGA TTTGGAGTCT AATAGAAATG CCAAGTCATT CTCTGGGTTT  
 2821 GGGCCCGCCC CAACTGTCAT TGCTTGTAGA AATTGGTAAC CCAATTTCTC TATGTTGGGG  
 2881 CCCCACCCCA ACTTACATTG T

10 SEQ ID No. 107

*S aureus strain MW2 (SSL 1 -11)*

15 139021 taaaaattat gaaattttaa gogatagcaa aagcaagttt agcattggga atgttagcaa  
 139081 caggtgtaat tacatcgaat gtacaatcag tacaagcgaa aacagaagtt aaacaacaaa  
 139141 gtgaggctga tttaaaactt tattataatg gaccaagttt tgaatataaa aaagtaactg  
 139201 gatattgatt tattgaaggt aaagatagat ttattgattt tatatacaat ggacaatata  
 139261 ataaaaatc tttagttggt tctgataaag ataaaataaa tgaagaagtt aaccagata  
 20 139321 tagatgtgtt tgtcgttaga gaaggaaacg gtagacaagc tgataatcat tcgattggtg  
 139381 gcataacaaa aactaataga ggagtgtatt atgactatat acacacacca atccttgaaa  
 139441 tcaagaaagg taaagaagaa ccacaaagta gtctatacca aatttataaa gaagacatct  
 139501 cactaaaaga acttgatttt aaattaagaa agcaattaat tagtcaaagt gccttgatt  
 139561 caaatggtct taaacaaggt caaattacaa ttacaatgaa tgatggcaca acacatacaa  
 25 139621 tcgatttaag tcaaaaactt gaaaaagaac gtatggcgca gtctatcgat ggcagacaaa  
 139681 tacaanaaat tctagtagaa atgaaataat actttctaac aacaaagcgc tatgttgaa  
 139741 agtgcttggt atggaaatat atggaagtta agcgcgtac tgttgcttag ctctctttt  
 139801 tgaggggaaa agttacaaaa ctacacaaaa cagtcgcacc acgcattatg ttttgcttaa  
 139861 atagcttaat catattttat gaatagttaa aaacaggtta atgtgaatat ctgaatacag  
 30 139921 ctctataat atgggtgtat ggttcaaatt acgtaataaa acaatctaata tataatagat  
 139981 tggagcatat agctatgaaa atgaaatcaa ttgtaaaaat aagtttggtt ttaggaatat  
 140041 tagcaacagg tgtaaacact acaacggaaa aaccagttca tgccgaaaag aaacctattg  
 140101 taataagtga aaatagcaaa aaattaaaag cttattatac tcaacctagt attgaatata  
 140161 aaaaatgtgac aggttatatc agtttctatt aaccaagtat taaattttatg ttttgcttag  
 35 140221 atggtaattc tgtaataaat attgctttta ttggcaaaga taagcaacat tatcatcagg  
 140281 gtgtacatcg taatcttaat atatttttac ttaatgagga taagagattt gaaggtgcaa  
 140341 agtactccat tggcgggtatc acgagtgcaa acgataaagc tgtcgaccta atagcagaag  
 140401 caagagttat taaagcagat catattgggtg aatatgatta tgactttttc ccatttataaa  
 140461 tagataaaga agcaatgtca ttgattttta attaaagaaa taccctattg  
 40 140521 ataattatgg tctttacggt gaaatgagta cagggaaaat taccgtcaaa aagaaatact  
 140581 acggaaagta tacatttgaa ttggataaaa agttacaaga agaccggatg tccgatgtta  
 140641 tcaatgtcac agatattgat agaattgaaa tcaaaagttag aaaagcataa cacacatact  
 140701 tgacgacgaa ataatttgaa attgaaatag agaggttaag tgacgatcaa acgttgctta  
 140761 acttcttttt aatgctttaa aatcatttca aaggcacata gaaacgctat attaacctca  
 45 140821 taatcactca ttattttttg cttaaattac ttaataatac ttcaataatt gttaaaagg  
 140881 gtttaattgt attatcttag aatgccatct ataattgatg tgtatgattc aaattacgta  
 140941 aaaagacaat cgaatataat atagattgga gcatacaatt atgaaaatga gaacaattgc  
 141001 taaaaccagt ttagcactag ggcttttaac aacaggcgca attacagtaa cgacgcaatc  
 141061 ggtcaaagca gaaaaaatac aatcaactaa agttgacaaa gtaccaacgc ttaaagcaga  
 50 141121 gcgattagca atgataaaca taacagcagg tgcaaatcca gcgacaacac aagcagctaa  
 141181 cacaagacaa gaacgcacgc ctaaaactga aaaggcacca aataactaat agggaaaaaac  
 141241 ctacagcttc aaaatagaaa aaatatcaca acctaataca gaagagcaga aaacgcttaa  
 141301 tatatcagca acgcccagcg ctaaaacaaga acaatcaca aacgacaaccg aatctacaac  
 141361 gcagcaaat aaaatgacaa cacctccatc acaaaacacg ccacaaccaa tgcaatctac  
 55 141421 taaatcagac acaccacaat ctccaacctt aaaaacaagc caaacagata tgactoctaa  
 141481 atatgaagat ttaagagcgt attacacgaa accgagtttt gaatttgaaa agcagtttg  
 141541 atttttgctc aaaccttgga cgacggttag gtttatgaat gttattccaa ataggttcat  
 141601 ctataaaaata gcttttagttg gaaaagatga gaaaaaatat aaagatggac cttacgataa  
 141661 tatcgatgta tttatcgttt tagaagacaa taaatatcaa ttgaaaaaat attctgtcgg  
 60 141721 tggcatcagc aagactaata gttaaaaagt ttaatacaaa gtagaattaa gcattactaa  
 141781 aaaagataat caaggtatga tttcacgcga tgtttcagaa tacatgatta ctacgaaga  
 141841 gatttccttg aaagagcttg atttttaatt gagaaaacaa cttattgaaa aacataatct  
 141901 ttacggtaac atgggttcag gaacaatcgt tattaataatg aaaaacggtg ggaaatatac  
 141961 gtttgaaata caaaaaaac tgcaagagca tcgtatggca gacgtcatag atggcactaa  
 65 142021 atttgataac attgaagtga atataaaaata atcatgacat tctctaaata gaagctgtca  
 142081 tcggaaaaaac aagaagttaa gtgacaacgg tttacatggt gcttagcttc tttattatg  
 142141 cgtaattgatg taaaaagacg aatattcatt tgtttgtaaa agtggcattt ctatgtctta  
 142201 aaagtgcaga aacttcaaat gtgccaaagt ttgaatcaca tcaaatcat ttttatttaa

	142261	cgaacattat	ggattttctta	atctacttta	cgatgattca	aatatagtta	aacaagggttt
	142321	aattgtgaatg	gagcaatgcg	ccatctataa	taaagctgta	tgattcaatg	aattgtaatcg
	142381	aacaaatcta	ataattacga	atggagcata	caactatgaa	aataacaaca	attgctaataa
5	142441	caagtttagc	actaggcctt	ttaacaacag	gtgtaatcac	aacgacaacg	caagcagcaa
	142501	atgcgacaac	accgccttca	actaaagtgg	aaacaccgca	acaagtagca	aattgcaacaa
	142561	caccatcttc	aactaaagtg	gaagcaccgc	aacaagcagc	aaacgcgaca	acaccatctt
	142621	caactaaagt	agaagcaccg	caatcaaaac	caaacgcgac	aacaccatct	tcaactaaag
	142681	tggaagcacc	gcaacaagca	gcaaacgcga	caacaccacc	ttcgtctaata	gtagacacat
10	142741	caccaccaca	atcgccaacc	acaaaacaag	taccaacaga	aataaatcct	aaatttaaaag
	142801	attttaagagc	gtattatacg	aaaccaagtt	tagaatttaa	aaatgagatt	ggtattatctt
	142861	taaaaaaatg	gacgacaata	agattttatga	atggtgtccc	agattatctt	atatataaaa
	142921	ttgcttttagt	tggtaaagat	gataaaaaat	atggtgaagg	agtacatagg	aattgtcgatg
	142981	tattttgtcgt	tttagaagaa	aataattaca	atctcgaaaa	atattctgtc	ggtggtatca
15	143041	caaagagtaa	tagtaaaaaa	gttgatcaca	aagcaggagt	aagaattact	aaggaagata
	143101	ataaagggtac	aattctctcat	gatgtttcag	aattcaagat	tactaaagaa	cagatttcctt
	143161	tgaaagaact	tgatttttaa	ttgagaaaaa	aacttattga	aaaaaataat	ctgtacggta
	143221	acgttgggttc	aggtaaaatt	gttattaaaa	tgaaaaacgg	tggaaggtac	acgtttgaat
	143281	tgacacaaaa	attacaagaa	aatcgcatgg	cagatgtcat	agatggcact	aattattgata
20	143341	acattgaagt	gaatataaaa	taattcatgac	attctctaaa	tagaagctgt	catcggaataa
	143401	acaagaagtt	aagtgcacac	ggcctacatg	ttgcttagct	tcttttggtta	tggtcgatga
	143461	tttgagaacc	cgaattttcg	atgggtccaa	atatgacgtg	gaagagtcct	gaatttatctt
	143521	gtaaatccct	atctatcggg	tgtggagcac	aacggaatca	gttttatttta	acgaacatta
	143581	tagattcctt	aatttactta	ataatgattc	aattgattatt	aaacatgggt	taattgtgaaa
25	143641	ggtcaaatat	gccaatata	ataaagctgt	atgattcaat	agacgtaagc	gaacaaatct
	143701	aataattacg	aattggagcat	acaactatga	aaatggcagc	aattgcgaaa	gcaagtttag
	143761	cattaggtat	tttagcaaca	ggaacaataa	cgtcattgca	tcaaaactga	aattgcgagt
	143821	aacatgaagc	aaaatatgaa	aattgtgaca	aagatatctt	tgacttaaga	gattactata
	143881	gtggcgcaag	taaggaaact	aaaaatgtta	ctgggttatcg	ttatagcaaa	ggtggcaagc
30	143941	attaccttat	ctttgataaa	catcaaaagt	tcactagaat	acaaattttt	ggtaaagata
	144001	tagaaagatt	taaagcacgc	aaaaatccgg	gattagacat	atgtgtgtgt	aaagaagcgg
	144061	aaaaccgtaa	tgccacagtg	ttttcatatg	gtgggtgtcac	taagaaaaat	caagacgctt
	144121	attatgatta	tataaacgca	ccaagatttc	aaatcaagag	agatgaaggt	gacggatttg
	144181	ctacgtacgg	tagagtacac	tacatttata	aagaagagat	ttcacttaaa	gaactcgact
35	144241	ttaaattgag	acagtattta	attcaaaatt	ttgatctgta	taaaaagttt	cctaaagata
	144301	gtaagataaa	agtataaatg	aaagatggcg	gctattatac	gtttgaactt	aataaaaaat
	144361	tacaaacaaa	tcgcatgagt	gacgtcattg	acggtagaaa	tattgaaaaa	atagaagcca
	144421	acattagata	atgcaatgaa	atatggataa	tagtaaaaata	tggtatgtat	agaggattta
	144481	ggcaacataa	gttgcttagc	ttcttttttg	tggttgagag	atgaaaaaga	agcgtatcga
40	144541	tgaataataa	aaacaccaat	aaaacttggt	ggaatagttg	atacctatag	tcgcgcgttg
	144601	tccttttctg	gacatgaaac	aattgtggaaa	acataattaa	attgagggaa	agtgtgaata
	144661	gttaaaaaag	ctgcgttaag	tttaaaaaat	agattaacgc	tggttaggatt	ctattaatata
	144721	gcttaacatt	ggttcaaaaa	tagttaaaaa	gaggttaatt	catagccttg	tattccgctt
	144781	atataatgat	agtagattgt	tcgtattacg	taattgaaat	aattcatataa	aaatatatta
45	144841	agacaaaatt	tataaataga	ttgggagaat	agtactatga	aattaaaagc	ggttagctaaa
	144901	gcaacattag	tattgggatt	gttagctact	ggtgtaataa	caacagaaag	tcaaacagta
	144961	aaagcggcag	aatcaactca	agggtcaacac	aattataaat	cattaaaaata	ctactatagc
	145021	aaagcgaata	tagagttaat	aaatgtatag	ggtctgtata	gacaacattt	aactgataaa
50	145081	ggtgcatatg	tatggaaaaa	tcttaaagat	tattatattg	ggctactagg	tgaagatagt
	145141	aagaaattca	aatcagatgt	atacggggac	ctagatgcat	tttttagtcat	agaagaagaa
	145201	cctgtttaaag	gaagacaata	ttcaattggc	ggtataagta	agacaaaatg	taaagaattt
	145261	aaagaaagag	aagtcgatgt	taaagtaaca	agaaaagcag	acagagatac	tacatcaact
55	145321	aaagatagta	aattttaaatt	tacaaaagaa	gaaatctcgt	taaaaagagt	agatttttaa
	145381	ttaagacaaa	aattgatgaa	agaagagaat	ttatacgtatg	caattaaacca	tagaaaaaggt
	145441	aaaattgttag	ttaaaatgga	agatgataag	ttttatactt	tcgaacttac	aaaaaaatta
	145501	caaccgcatac	gcatgggtga	cacgatagat	ggtacaaaaa	tcaaaagaaat	taattgttag
	145561	ctagaatata	aataatcttt	ggacaagcag	actagtaatt	gtagggaagt	taagcgataa
60	145621	catattgctt	agcttctttt	ttgttttgtt	atgatgaaaa	aaggagcggg	tttatgatca
	145681	agtttttggg	aaaacgggtg	atacttatag	tcgcgcgttg	tccttttctg	gacatgaaac
	145741	aattgtggaaa	acataattaa	attgagggaa	agtggtgaata	gttaaaaaat	tagtattgtg
	145801	ttataaaaaa	taattaatac	tggttaggatt	tcattaacta	acttaacggt	ggttcaaaaa
	145861	tagttaaaaa	gaggttaatt	catagcgcag	tatctcactt	atataatgat	agtagattgt
	145921	tcgtattacg	taattgaatt	aatcatataa	aaatatatta	agacaaaatt	tataaataga
	145981	ttgggagaat	agtactgtga	aattaaaaac	ggttagctaaa	gcaacattgg	cataggcctt
65	146041	attaactact	ggtgtgatta	catcagaagg	ccaagcagtg	caagcaaaag	aaaagcaaga
	146101	gagagtacaa	catttatatg	atattaaaga	cttacatcga	tactactcat	cagaaagtgt

	146161	tgaattcagt	aatattagtg	gtaagggtga	aaattataac	ggttctaacg	ttgtacgctt
	146221	taaccaagaa	aatcaaaatc	accaattatt	cttatcagga	aaagataaag	ataaatataa
	146281	agaaggcctt	gaaggccaga	atgtctttgt	ggtaaaagaa	ttaattgatc	caaacggtag
5	146341	actatctact	gttggtggtg	taacgaagaa	aaataaccaa	tottctgaaa	ctaatacacc
	146401	tttattttata	aaaaaagtgt	atggcgga	tttagatgca	tcaattgaat	cattttta
	146461	taataaagaa	gaagtttcac	tgaagaact	tgatttcaaa	attagacaac	atttagttaa
	146521	aaattatggt	ttatataaag	gtacgactaa	atacggtaag	atcactttca	atttgaaaga
	146581	tggagaaaag	caagaaattg	atttaggtga	taaattgcaa	ttcgagcaca	tgggcgatgt
	146641	gttgaatagt	aaggatattc	aaaaatagc	agtgactatt	aatcaaat	aaagtaagta
10	146701	atcaatgact	ctaaagtaat	aaatttgaag	cagcttagcg	atgaaatg	gaatagatac
	146761	gtacacctta	cataaaggag	cgtattttaa	acaacctgt	cgtaggctt	ttttacg
	146821	ttataacgca	gggtatgagc	gtactaaaa	ttcacattac	ttctgaaagt	gatgtccatt
	146881	gaatattta	tagttcttca	tttaaccatga	tttaatttta	attaaacgag	tgtaaatgtc
	146941	agtctgtctc	aatgcccttt	ataataaatg	tgtattattc	aaattacgta	ataaaagcaa
15	147001	tccaatatat	taagattgga	gcatatgaat	atgaaattta	cagcgatagc	taaagcgata
	147061	tttgtattag	gaatattta	aaacaagtga	atgataacag	aaaatcaatc	ggttaatgca
	147121	aaaggaaagt	atgaaaaaat	gaaccgttta	tatgatacaa	acaagttaca	tcaatactat
	147181	tcaggaccta	gttatgagtt	aacaaatgtt	agtggccaaa	gtcaagggtta	ttatgactct
	147241	aacgttttgc	tttttaacca	acaaaaatcaa	aagttccaag	tgtttttatt	gggaaaagat
20	147301	gaaaataaat	acaaagaaaa	aacacatggt	ttagatgtct	ttgcggtacc	ggaattagta
	147361	gatttagatg	gaagaatatt	tagtgtagt	ggtgtaacaa	agaaaaatgt	aaaatcaata
	147421	tttgagtctc	taagaacgcc	gaacttacta	gttaaaaaaa	tagacgataa	agacggtttt
	147481	tcgtatgatg	aatttttctt	tattcaaaaag	gaagaagtat	cattgaagga	acttgatttc
	147541	aaaataagaa	aactgtta	taaaaaatac	aaattgtatg	aaggggcagc	tgataaaggt
25	147601	agaattgtta	ttaatatgaa	agatgaaaat	aagtatgaaa	ttgatttaag	tgataaatta
	147661	ggtttcgagc	gtatggcaga	tgctattaat	agtgaacaaa	ttaaaaacat	cgaagtgaat
	147721	ttgaaataat	caatgatata	tatagaatga	aagcttaaga	agcgggttaa	taatcccatg
	147781	tttaattgatt	ttgatacgtg	ttttaataat	aaaaacatat	cgaacattga	ctacgttatt
	147841	aagctgcttt	ttgtacact	ttgtatcgaa	taacttaaga	tctaaaaacta	atcggaaga
30	147901	acaatgattc	ccctaaaaaa	atattatgtt	ctattaaaaa	tcagttaata	cgaatgttaa
	147961	catacgtttg	attttcatta	ataatgattc	aagtttattt	aatgagcgt	taatgtcagt
	148021	ctggtttgat	gcaccttata	ataaagacag	atagttcaaa	ttacgtaata	atacaatcc
	148081	aatatatcaa	gattggagca	aataaatatg	aaattttacag	cattagcaaa	acaacatta
	148141	gcattaggaa	tattaactac	aggtgtgttt	acaacagaaa	gtaaagctgt	tcacgcgaaa
35	148201	gtagaacttg	atgagacaca	acgcaaatat	tatatcaata	tgctacatca	atactattct
	148261	gaagaaagt	ttgaaccaac	aaatattagt	gttaaaagcg	aagattacta	tggctctaac
	148321	gttttaaa	ttaaacaacg	aaataaagct	tttaaagtat	ttttacttgg	tgacgataaa
	148381	aataaatata	aagaaaaaac	acatggcctt	gatgtctttg	cagtacctga	attaatgat
	148441	ataaaaagtg	gcataatag	cggtggcggt	ataacaaaga	aaaatgtgag	atcagtgttt
40	148501	ggatttgtaa	gtaatccaag	tctacaagtt	aaaaaaatcg	atcctaaaca	tggcttttcg
	148561	ataaatgagt	tgttctttat	tcaaaaggaa	gaagtatcgt	tgaaggaact	ggatttttaa
	148621	ataagaaaaa	tgtagtcga	aaaatataga	ttgtataaag	gcgcgtcaga	taaaggtaga
	148681	atcgttatta	atatgaaaga	cgaaaagaaa	tatgtaattg	atgtaattga	aaaatgaagt
	148741	tttgatcgta	tggttgatgt	aatggatagt	aagcaaatga	aaaatattga	agtgaatttg
45	148801	aatttaattaa	gtataataac	ttaagaagcg	acttaacgac	aaaatgtgaa	ttgacatgca
	148861	tgctcttaaa	taagggaactg	tggttaaatc	attactgttg	ttaggtgtt	ttttgcgttt
	148921	caaagagcag	aacagagtaa	catcatcagt	tgtagtaaac	gataatctag	taaaacaact
	148981	aaatgaaata	atgaaattca	tttaacctga	acattaaaaa	atatttgttt	ttcattaaga
	149041	ataattcaag	tatattttaa	tcgagggttaa	ttatcgtatg	aaacgatgca	cgttataata
50	149101	aaaatgtatg	attcaaatga	cgtaattgaaa	acaatccaat	atattaagat	tggagcaaaa
	149161	caatatgaaa	ttaacagcga	tagctaaagc	tgcattagct	ttaggaattt	taacaacagg
	149221	aactttaaca	acagaagttc	atccaggtca	tgcaaaaaca	aatcaaaagt	cagtaaaata
	149281	acatgacaag	gaagcattat	accgatacta	cactggaaaag	actatggaaa	tgaaaaatat
	149341	tagtgctttg	aaacatggta	aaaataactt	gcgttttaag	tttagaggta	ttaagattca
55	149401	agtttttactg	cctggaaatg	ataaaagtaa	atctcaacag	cgtagttatg	aggggttaga
	149461	tgtgtttttt	gttcaagaaa	aaagagataa	gcacgatata	ttttatactg	ttggtggtgt
	149521	aatacagaat	aataaaacat	ctggagttgt	cagtgaccca	atattaaata	tttcaaaaga
	149581	aaagggtgaa	gatgcttttg	tgaagggtta	cccttattac	attaaaaaag	aaaaataaac
	149641	actaaaagag	ctggattata	agttgagaaa	gcacttaatc	gaaaaaatatg	gactttataa
60	149701	aacaatctca	aaagatggta	gggtcaaatat	tagcttgaaa	gatggcagtt	tttataacct
	149761	tgatttaaga	tctaaattaa	aattcaaatat	tatgggggaa	gtcatagaaa	gcaaacaaat
	149821	taaagatatt	gaagttaact	taaagtaaat	aattacgaat	aataaaaagt	aattgaagcg
	149881	gcttaacgat	gaaaagtaaa	ttgatgcgca	taccttacca	aaaggatgca	tcaatcgata
	149941	tcgtcggttaa	gctgttttgg	tttacgtttc	atggattcta	ccccaat	cataaatata
65	150001	aaaattccac	caccaacatc	aaaattctca	acatcgcaac	atacccaaat	gttataataa

	150061	atctattaca	caaagagata	aattactttat	tcaaaaggcgg	aggaatcaca	tgtctattac
	150121	tgaaaaacaa	cgtcagcaac	aagctgaatt	acataaaaaa	ttatggtcga	ttgcgaatga
	150181	tttaagaggg	aacatggatg	cgagtgaatt	cogtaattac	attttaggct	tgattttcta
	150241	tcgcttctta	tccgaaaaag	cagaacaaga	atatgcagat	gcgttggcag	gtgaagatat
5	150301	cacgtatcaa	gaggcatggg	cagatgaaga	atatcgtgaa	gacttaaaag	ctgaattaat
	150361	tgatcaagtc	ggttacttca	ttgaaccaca	agattttatc	agtgcgatga	ttcgtgaaat
	150421	tgaaacgcaa	gatttcgata	tcgaacatct	ggcgacggca	attcgtaaaag	ttgaaacatc
	150481	aacattaggt	gaagaaagt	aaaatgactt	tatcggactg	ttcagcgata	tggacttaag
	150541	ttcaacgcga	ctaggttaaca	atgtcaaaga	acgtactgct	ttaatctcta	aagtcattggt
10	150601	taatcttgac	gattttaccat	ttgttcacag	tgatatggaa	attgatattgt	taggtgatgc
	150661	atatgaattc	ctaattgggc	gctttgcggc	gacagcaggt	aaaaaagcag	gogagttcta
	150721	tacaccacaa	caagtatcta	agatactggc	gaagattgtc	acagacggta	aagataaatt
	150781	acgtcatgtg	tacgacccaa	catgtgggtc	cgttccatta	ttgttacgtg	ttggtaaaga
	150841	aacgcgaagt	tatcgttatt	tcgggtcaaga	acgttaacaat	accacttaca	acttagcacg
15	150901	catgaacatg	ttattacatg	atgtacgtta	tgaaaatttc	gatattccga	atgatgacac
	150961	gttggaaaaa	ccagcctttt	taggacatac	atttgatgcg	gttattgcga	accaccata
	151021	cagtgcgaaa	tggacagcag	attcaaaaatt	tgaaaatgac	gaacgattca	gcggatacgg
	151081	caaacttgcg	ccaaagtcca	aagcagactt	tgccctttatt	caacacatgg	tacattactt
	151141	agacgatgaa	ggtaccatgg	cogttgtact	cccacatggg	gtcttattcc	gtgggtgctgc
20	151201	agaaggtgtc	attcgtcgtt	atttaattga	agaaaagaac	tacttagaag	ccgtgattgg
	151261	cttaccagcc	aatattttct	atgggacaag	tattccaaca	tgtattttag	tatttaaaaa
	151321	atgtcgcgca	caagaagact	atgtattatt	tatcgatgca	tccaatgatt	ttgaaaaagg
	151381	aaaaaatcaa	aaccatttta	ccgatgccca	agtcgaacgc	attattaaca	catataagcg
	151441	taaggaaaca	attgataaat	atagctacag	tgcgacatta	caagagatcg	ccgataacga
25	151501	ttacaactta	aacataccga	gatattgtcga	tacattcgaa	gaagaagcac	cgatttgattt
	151561	agatcaagtc	caacaagatt	tgaaaaatat	cgacaagaa	atcgcaaga	ttgaacaaga
	151621	aatcaatgca	tacotgaaag	aacttggggg	gttgaaagat	gagtaataca	caaaaagaaa
	151681	atgtgccaga	gttgagggttc	ccagggtttg	aaggcgaatg	ggaagagaag	aagttagggg
	151741	accttactac	caaaataggt	agtggaaaga	ctcccaaagg	tggaaagtga	aactatacaa
30	151801	acaaaggcat	accattttta	aggagtcaaa	atattagaaa	tggtaaatta	aatcttaatg
	151861	acttagttaa	tatttagtaaa	gatatagatg	atgagatgaa	aaatagtaga	acgtactatg
	151921	gtgatgttct	tttaaatatt	acaggagcat	caataggtag	aacagccatt	aatctgatag
	151981	ttgaaatata	tgctaattta	aatcaacatg	tatgtattat	tagattgaaa	aaagagtatt
	152041	attataattt	ttttggacag	tatctattat	caagaaaagg	taaaaggaaa	attttccttg
35	152101	cacaaagtgg	aggtagtcga	gaaggactaa	acttcaaaga	aattgcta	ttaaaaatct
	152161	tcaccccaac	tatatattgaa	gagcagcaaa	aaataggcga	attcatcagc	aaacttgacc
	152221	gacaaattga	attagaagaa	caaaaacttg	aattacttca	gcaacagaaa	aaaggctata
	152281	tgacagaaaat	cttctcgcaa	gaattgcgat	tcaaagatga	ggaaggtaaa	gattatccag
	152341	attggaaatc	aaaatcaatt	caagaaatat	ttgagaataa	gggtggcact	gctctagaaa
40	152401	cagaatttta	ttttgacggt	aattataaag	ttataagtat	aggaagttat	tctataaata
	152461	gcacttataa	tgatcaaaaat	ataagagtca	ataaaaaata	aaaaactgaa	aaatatattt
	152521	tatcaaaagg	cgacttagca	atggatttaa	atgataaaac	aaaagatggg	aaaattatag
	152581	gtagaagtat	atttatagat	aaagataatc	aatatattta	taatcaaga	actgaaagat
	152641	taataccatt	tgctgaaaat	gataataaat	ttttatgggt	cttaatgaat	acagatttaa
45	152701	ttagaaataa	aataaaagggt	atgatgcaag	gagcaaccca	agtttatata	aattattcat
	152761	ctattaaatt	gatattctata	caattgccac	ttcttgaaga	acaacagaaa	ataagagggt
	152821	ttctagaagt	tttatctgga	ataactacta	aacaattgca	caagatagac	caattaaaag
	152881	agaggaaaaa	ggcgttttta	cagaaaaatgt	ttatttgatt	tgctcgctgca	atatagtttt
	152941	tattattttgt	ttatttcaga	tgtttcacca	tcatattgcg	taactttttac	aaataagaaa
50	153001	taaagttcaa	tgaaatcaaa	aacgaacatt	aaatttaggc	actgtgatag	cacagtgctc
	153061	tttttgtgtc	gaaattgtgt	acagaataag	tagttaaata	aagattaagt	tgagataaag
	153121	tgtttattcgt	aaataaaaaga	gagtagatcg	ataggaattg	aatgatatta	gttaactatt
	153181	tattaaatta	cttaataatg	attaaattttt	agttaaaagta	agtttaattgt	gaagcacgac
	153241	cattgctcat	tataatgaat	gaggattgtt	cgtattgctg	aatagaataa	atcaaataga
55	153301	ctaaaaattg	ggagcataga	attatgaaat	taaaaaatat	tgctaaagca	agtttagcac
	153361	tagggattttt	aacaacaggg	atgattacaa	ctactgctca	gccagtaaaa	gcaattgagc
	153421	aaagcagatt	atcagttact	tcaaaagata	cacaagaatt	aaaaaaatc	tacagtggaa
	153481	caggatataa	ttttcaaat	gtgagtgggt	atagagaagg	taataaaatg	aactattttg
	153541	atggaccaca	acttaagtga	gttactttac	ttggcacaga	caaagaaagg	tttaaggacg
60	153601	atgaagatta	tgaaggactt	gatgtatttg	ttgtaagaga	agggtcaggt	aaacacgcag
	153661	ataatatatc	aattgggtgga	attacaaaaa	caaataagaa	tcaatataaa	gaccctgtac
	153721	aaaacgttaa	tttattgact	tctaagagta	acggtcaaaa	tactgcttct	gtgacttcag
	153781	ataactatag	catcaataaa	gaagaaattt	cattaaaaga	acttgatttc	aaactaagaa
	153841	agcaattaat	tgataaacat	gatctttata	agacagagcc	taaagacagc	aaaattaaag
65	153901	tttctatgaa	aaatggcggc	tactatacgt	ttgaattaaa	taaaaaatta	cagcctcatc

	153961	gcatgggtga	tacgattgat	agtagaaata	taaagaaaat	tgaagtgaat	ttataataat
	154021	attcgagggga	gtatatcatg	agagaaaatt	ttaaattacg	taaaatgaaa	gtcgggttag
	154081	tatcagttgc	aattactatg	ttatatatca	tgacaaatgg	cgaagcagaa	gcatctgaag
5	154141	gtagccaaac	tgtaaagaac	ccaaagggtga	atgcaactga	agaaatcaaa	gttggatcac
	154201	aaccagttca	aaataatcaa	gaagtaagtt	cggaacaaac	aaaaaagaat	tttggttaatt
	154261	tagaccccat	taaacctggt	gctcaaaaagg	taacagggac	tactttaccc	agccatatta
	154321	ttctaatagaa	tatagatggt	aaaagtgcgtg	atccagtaga	tggaggaaaat	agtgatttag
	154381	tatttgctga	tgaaaacgga	agattcgagt	atccactaca	caatagaaaa	attgttcata
	154441	atcaagaaat	tgagggttctg	tcatccagtc	ctgatttagg	tgatgatgaa	gaagatgaag
10	154501	aagtagaaga	agattcaact	gaaaaagctg	gtactgagga	agaaaacaca	gatgctaaag
	154561	ctacatacac	aacaccacga	tatgaaaaag	cgtatgaaat	accgaaagaa	caactaaaag
	154621	aaaaagatgg	acatcaccaa	gttttcatcg	aacctattac	tgaagggttca	ggtattatta
	154681	aagggtcatag	gtctgtaaaa	ggtaaagttg	ctctatctat	taataataaa	tttattaatt
	154741	ttgaagaaag	agctaaagat	ggaattagta	agaagatac	taaagctagt	tccgatgggtg
15	154801	tttggatgcc	tattaatgaa	aaaggatatt	ttgattttga	tttcaaaaaa	aatccttttg
	154861	ataaacttaga	gttaaagaaa	aatgatgaaa	tctcattaac	atttgcacct	gatgatgaag
	154921	atgaagcatt	gaagtcatta	attttcaaaa	ctaaagtaac	gagtttagaa	gatattgata
	154981	aagcagaaac	taaatatgac	cataactaaag	tggaaaaaagt	aaaagtattg	aaagatgtta
	155041	aagaagattt	acatgtagat	gaaatttatg	gaagtttgta	ccatacagaa	caaggtaaag
20	155101	gtattctcga	taaacaggga	actaaagaaa	ttacaggtaa	gactaaattc	gcgaatgcag
	155161	tagtgaaaagt	atattctgac	ttagggtgatg	cgcaactggt	tcctgatatt	caagtagatg
	155221	aaaatggtaa	atthagcttt	gatgctgaaa	aagctgggtt	cagattacaa	aatggagaaa
	155281	cattgaattt	tgcagtagtt	aaacctatta	ctgggtgagct	attacatcaa	ggattcggtt
	155341	ctaagtatat	cgatgtttat	gaatctccgg	aagaaaagaa	agaacgtgaa	tttgaagaga
25	155401	aacttgaaaa	cacgcctgca	tatcataaat	tacatgggtga	taaaattgtc	ggctatgatg
	155461	ttcaaggtaa	tccatcaact	tggttctatc	cattaggtga	aaagaaagtt	gaacgtaagg

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
6 October 2005 (06.10.2005)

PCT

(10) International Publication Number  
**WO 2005/092918 A3**

(51) International Patent Classification:

*C07K 14/31* (2006.01)      *C07K 19/00* (2006.01)  
*A61K 39/085* (2006.01)      *C12N 5/06* (2006.01)  
*A61K 48/00* (2006.01)      *C12N 7/01* (2006.01)  
*C12N 15/62* (2006.01)

(21) International Application Number:

PCT/GB2005/001084

(22) International Filing Date: 22 March 2005 (22.03.2005)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

0406408.5      22 March 2004 (22.03.2004)      GB  
0412419.4      3 June 2004 (03.06.2004)      GB

(71) Applicant (for all designated States except US): **UNIVERSITY COLLEGE LONDON** [GB/GB]; Gower Street, London WC1E 9BT (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CHAIN, Benjamin, Michael** [GB/GB]; 46 Cleveland Street, London W1T 4JF (GB). **NAIR, Sean, Paddy** [GB/GB]; 256 Grays Inn Road, London, WC1X 8LD (GB).

(74) Agent: **WOODS, Geoffrey, Corlett**; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:

27 April 2006

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TARGETING POLYPEPTIDE

(57) Abstract: A targeting polypeptide is provided that may be used to target a chosen antigen to an antigen presenting cell. Complexes comprising such targeting polypeptide and antigen, nucleic acids and vectors encoding them, and cells comprising the nucleic acids and vectors may be used in methods of immunisation and enhance the immunogenicity of the antigen.



WO 2005/092918 A3

# INTERNATIONAL SEARCH REPORT

ational Application No  
/GB2005/001084

## A. CLASSIFICATION OF SUBJECT MATTER

C07K14/31 A61K39/085 A61K48/00 C12N15/62 C07K19/00  
C12N5/06 C12N7/01

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K A61K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS, EMBASE

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>US 2002/177551 A1 (DAVID S. TERMAN) 28 November 2002 (2002-11-28) page 2, paragraphs 14,19 page 3, paragraphs 21,22 page 8, paragraph 51-57 page 12, paragraph 89 - page 16, paragraph 132 page 21, paragraph 180 - paragraph 182; example 3</p> <p style="text-align: center;">----- -/--</p>	1-25

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*G\* document member of the same patent family

Date of the actual completion of the international search

20 February 2006

Date of mailing of the international search report

09/03/2006

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Montero Lopez, B

## INTERNATIONAL SEARCH REPORT

national Application No  
T/GB2005/001084

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SHIMIZU MOTOMU ET AL: "A novel method for modification of tumor cells with bacterial superantigen with a heterobifunctional cross-linking agent in immunotherapy of cancer." MOLECULAR BIOTECHNOLOGY, vol. 25, no. 1, September 2003 (2003-09), pages 89-94, XP008052170 ISSN: 1073-6085 abstract page 89, right-hand column, paragraph 2 - page 90, right-hand column, paragraph 1 -----	1-25
A	ERIC MURAILLE ET AL.: "T cell-dependent maturation of dendritic cells in response to bacterial superantigens" JOURNAL OF IMMUNOLOGY, vol. 168, 2002, pages 4352-4360, XP002368601 abstract page 4355, left-hand column, paragraph 2 - right-hand column, paragraph 1 page 4357, left-hand column, paragraph 2 - page 4358, right-hand column, paragraph 1 -----	1-25
A	ALOUF J E ET AL: "Staphylococcal and streptococcal superantigens: molecular, biological and clinical aspects" INTERNATIONAL JOURNAL OF MEDICAL MICROBIOLOGY, URBAN UND FISCHER, DE, vol. 292, no. 7-8, 2003, pages 429-440, XP004959934 ISSN: 1438-4221 the whole document -----	1-25
A	WO 03/012111 A (LORANTIS LIMITED) 13 February 2003 (2003-02-13) page 2, last paragraph - page 4, line 15 page 8, line 14 - page 12, line 17 page 39, line 23 - page 40, line 10; example 1 -----	1-25

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/GB2005/001084

## Box II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
Although claim 17, as far as encompassing an in vivo method, and claim 24 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the composition.
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International Application No

/GB2005/001084

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 2002177551	A1	28-11-2002	NONE		
WO 03012111	A	13-02-2003	EP	1419256 A2	19-05-2004
			JP	2005524381 T	18-08-2005
			US	2004213797 A1	28-10-2004

